AD A076818

Research Memorandum 76-16

EVALUATION OF THE EFFECTIVENESS OF TRAINING DEVICES: ELABORATION AND APPLICATION OF THE PREDICTIVE MODEL

George R. Wheaton, Paul W. Fingerman,
Andrew M. Rose, and Russell L. Leonard, Jr.

American Institutes for Research

UNIT TRAINING AND EVALUATION SYSTEMS TECHNICAL AREA

U. S. Army

Research Institute for the Behavioral and Social Sciences

July 1976

79 11 13 288

DISPOSITION FORM

For use of this form, see AR 340-15, the proponent agency is TAGCEN.

REFERENCE OR OFFICE SYMBOL	SUBJECT
PERI-TP	Clearance and Transmittal of Reports to DTIC

TO DDC-DAA-1
ATTN: Mr Schrecengost

FROM ARI Rsch Pub Group

DATE 8 Nov 79
Ms Price/48913

CMT 1

1. The reports listed on Inclosure 1 are approved for public release with unlimited distribution (50 numbered ARI Research Memorandums, 74-1 thru 76-30).

2. These are among the previously unrecorded ARI reports which you identified to us 22 June 1979 as not in your retrieval system. The accompanying box contains at least one copy of each report for your retention and reproduction.

Helen S. Price

1 incl
List of reports, 1974-76

HELEN S. PRICE
Research Publications Group
Army Research Institute

A

DA FORM 1 FEB 62 2496 REPLACES DD FORM 96, WHICH IS OBSOLETE.

(14) ARI-RM-76-16

Unit Standards and
Performance Evaluation
Contract Number DAHC [illegible]

(11) Army Project Number
[illegible]

(9) Research ~~Memorandum~~ 76-16

(6) EVALUATION OF THE EFFECTIVENESS OF TRAINING DEVICES: ELABORATION AND APPLICATION OF THE PREDICTIVE MODEL.

(10) George R. Wheaton, Paul W. Fingerman,
Andrew M. Rose, [redacted] Russell L. Leonard, Jr.
~~American Institutes for Research~~

~~Angelo Mirabella, Work Unit Leader~~

~~Submitted by:~~
Frank J. Harris, Chief
Unit Training ~~and Evaluation~~ Systems Technical Area

(11) Jul [redacted]

(12) 85

Approved by:

Joseph Zeidner, Director
Organizations and Systems Research
Laboratory

J. E. Uhlaner, Technical Director
U.S. Army Research Institute for
the Behavioral and Social Sciences

408010 [illegible]

Research Memorandums are informal reports on technical research problems. Limited distribution is made, primarily to personnel engaged in research for the Army Research Institute.

EVALUATION OF THE EFFECTIVENESS OF TRAINING DEVICES: ELABORATION AND APPLICATION OF THE PREDICTIVE MODEL

CONTENTS

TABLES

FIGURES

1.0 INTRODUCTION

1.1 Background

One of the most complex problems facing Army planners is the design and development of effective training systems, particularly where the use of operational equipment for training purposes is impractical. A number of situations necessitate constraint in the use of operational equipment in a training role. These include reduced military budgets with consequent reduced availability of actual hardware for training purposes, reduced availability of large-scale training areas and ranges, and finally, growing concern for the ecological damage which can arise from mechanized field engagements.

In order to deal with the limitations resulting from the reduced use of operational hardware the Army has turned increasingly to the use of training devices which simulate the operational situation. These training devices are designed to meet the needs of a variety of students who enter the training situation lacking varying degrees of knowledge or skill. To the extent that exposure to the training device(s) imparts necessary knowledge and facilitates performance at specified criterion levels on an operational task, the training device is judged to be effective.

A basic problem is how to plan for, design, and develop a training system incorporating the different kinds of devices which will prove to be effective for a particular training and operational situation. The development starts with a statement of the requirement for training (e.g., a new weapon system is being developed for which trained operators will eventually be required; or, a training need is identified which cuts across several weapons systems). The next step is to identify what needs to be trained (the training deficits possessed by students who will be exposed to the training) and to specify the general training system which will meet these objectives. At this level the specification is still a "functional" one, oriented toward goals and objectives as opposed to hardware.

As planning of the training program progresses, the level of detail also increases and the "how" of training begins to be explored. Decisions are made regarding classroom versus on-the-job training, the length of the course(s), and the requirements for training devices and aids to support training. Many of these decisions are fairly straightforward and can be made by experienced training specialists familiar with the personnel needs and resources of the Army. Questions regarding training devices, however, are not readily amenable to such a process. How and when to use them, how to design them, and what to spend on them are issues which, in the past, have been dealt with in a fairly arbitrary manner due to the lack of objective bases on which to make such decisions. Sound instructional decisions regarding the use of training devices are contingent upon the development of a conceptual framework and methodology which can be employed to forecast training device effectiveness.

The primary goal of the present project is the development and evaluation of a model which can be used to predict and to evaluate the effectiveness of training devices. The modeling is particularly aimed at describing how device design, device use, training strategy, and individual ability interact to influence device effectiveness. Standards of effectiveness include both acquisition and transfer of military skills with emphasis placed upon transfer from training to operational settings. As described in the preceding reports on this project (Wheaton, Rose, Fingerman, Korotkin, and Holding, 1974; 1976), the first step in developing the model was to examine, and when possible, to build upon previous efforts. Toward this end, past methods and models dealing with the design or evaluation of training devices were reviewed, general theories of transfer were studied, and a host of specific variables were surveyed, particularly in terms of their impact on transfer. Based upon a synthesis of these inputs a preliminary model was formulated for predicting the effectiveness of a given training device at the various stages in its development. In the training-content by training-process model which emerged, device effectiveness was viewed as a function of: (1) the

potential for transfer, (2) the magnitude of the trainees' learning deficit, and (3) the appropriateness of the training techniques used to overcome that deficit. Subsequent efforts have sought to refine and clarify the model and to develop a plan for field testing and evaluation.

1.2 Purpose of the Report

As the second document in the series, the present report is concerned with the development of an updated form of the transfer of training model, and the synthesis of its components into a standardized method for its application. Major concerns are determination of the kinds of data necessary to apply the model, an assessment of the availability of the needed information, and the development of a feasible and reliable set of procedures for processing the data to generate predictions of potential training-device effectiveness.

The following sections of the report outline the steps followed to achieve these objectives. Section 2.0 outlines the rationale and development of the predictive model into its current form. In generating predictions about the training effectiveness of a given training device, the model combines data concerning the device's transfer potential, the learning deficits of the trainee population, and the extent to which the device incorporates various training principles and techniques which have been shown to have potential for enhancing training effectiveness.

Section 3.0 of the report describes the detailed procedures that are necessary to generate predictions from the model. Procedural issues include: the data which are needed, as well as how they are to be acquired and processed. The first procedural step involves the determination of transfer potential for the device in question. The model presents transfer potential as a joint function of: (1) task communality between the training device and the operational task, and (2) similarity between the device and the operational situation. The second step involves the determination of a Learning Deficit for the trainees who will be trained on the device. The third step requires an assessment of the extent to which the various training steps in the device utilize various principles

of training. The final procedural step in applying the model involves the process by which the data from the other steps are combined into an index or indices which reflect the device's effectiveness. The indicators relate to the percentage of training time saved by using the device as well as to various measures of trainee proficiency after a given amount of exposure to the device. The model also allows for an assessment of transfer that takes into account the type and amount of supporting classroom instruction which the trainees experience before practice on the device.

Section 3.0 also describes in detail the step-by-step procedures that the project staff followed in performing an experimental application of the model. Two training tasks used for this purpose were: (1) fire main gun (M60A tank) using the M-32 sight, and (2) apply burst-on-target adjustment of fire using the M-32 sight. Four devices were processed. These were the 17-4 Burst-on-Target Trainer, 17-B4 Burst-on-Target Trainer, M-55 Conduct-of-Fire Trainer, and SIMFIRE. These devices were selected in a manner designed to yield at least modest variation in their predicted efficacy for training the two criterion tasks so that the model's utility could be assessed across a range of predictions.

The purpose of this preliminary application was twofold. First, it was designed to assess the feasibility of applying the model. Feasibility includes an assessment of the time in a training device's "life cycle" at which the model can be applied, either at the Training Device Requirement (TDR) stage, or at the prototypic device stage. Another aspect of feasibility concerns the potential application of the model for systems versus non-systems training devices. The second purpose of the application was to determine the reliability of the procedures. In most cases, modeling data were collected independently from four senior project staff members and reliability indices were computed. These are reported in Section 3.0. This section also discusses the problems that the project staff encountered in applying the model and the resolutions which were

achieved. This information should prove invaluable in future applications and as guidance for Army personnel who may wish to apply the model.

Section 4.0 presents the development and derivation of a predictive equation using as parameters the outputs of the analyses discussed in Section 3.0. Section 5.0 consists of a discussion of the results, and suggestions for further research in the area.

2.0 THE MODEL

2.1 Background

The first report in this series (Wheaton, et al., 1976) presented a preliminary structural model for use in forecasting training device effectiveness. This preliminary model incorporated most of the central issues involved in training device effectiveness that were revealed through an analysis of previous models, methods, and empirical data. The model dealt with two major classes of variables:

1. Those associated with developing a training device which does, in fact, elicit the behaviors which are required in the operational situation; these were termed "Appropriateness" variables.
2. Those variables associated with actually learning these behaviors; these were called "Efficiency" variables.

Under "Appropriateness," the central issue was the transfer potential of the device. Assuming that the trainee became proficient on the tasks presented in the training situation, would he then meet the training requirements? To deal with this question, three types of analyses were proposed: (1) a communality analysis, (2) a criticality analysis, and (3) a similarity analysis.

In addressing the "Efficiency" issues, two major analyses were proposed. The first involved a determination of the trainee's learning deficit: an assessment of what trainees were actually required to learn. This was addressed by three proposed analyses: (1) determining whether appropriate skills and knowledge were already in the trainee's repertory; (2) establishing the proficiency requirements for the criterion transfer task; and (3) estimating how difficult it would be to learn the task. The second major analysis subsumed under "Efficiency" was the Training Techniques and Principles analysis. This proposed analysis was an attempt to make direct use of empirical data and training principles for a specific situation in order to depict the efficacy of training.

The basic input data for all of these proposed analyses was presumed to stem directly from or to be derivable from task analyses of the training and the operational situations. The preliminary model proposed two complementary schemes for conducting the task analyses, thereby determining the "Training Content": (1) a detailed task description, and (2) a behavioral taxonomic classification.

In summary, the preliminary model was represented as a training-content by training-process matrix. The content axis consisted of task analytic data, while the process axis was made up of two major headings, Appropriateness and Efficiency, and several subheadings. A functional model was only implied in the previous report; basically, it was assumed that the inputs to the Appropriateness and Efficiency analyses would be task or subtask descriptions (and the behavioral categories for these tasks) of the operational system and the training situation, combined with a physical description of the operational and training equipment. The precise nature of the measurements to be taken, the resultants of these individual analyses, and how these measures would be combined were unspecified in the preliminary model.

2.2 Current Structural and Functional Model

With minor exceptions, the model in its current state of development retains the basic structure of the preliminary model. While specific decisions regarding the implementation of the various analyses have been made, the basic rationale for the general types of analyses to be performed has remained unchanged. Training device effectiveness is still viewed as a function of the transfer potential of the device, the learning deficit of the trainers, and the extent to which appropriate training techniques are utilized in the device. As mentioned in the previous report, training effectiveness in general will be moderated by a host of potent variables external to the device itself, such as device acceptance, other instructional support, etc. While it is still felt that many of these variables would more appropriately be considered in a training system effectiveness model, provision has been made for an

"extended" application of the current device effectiveness model. For instance, it has been found that the type and amount of supporting classroom instruction can be incorporated into the learning-deficit portion of the model. Other examples of this extended model application will be presented below.

Figure 1 presents the structural and functional model at its present stage of development. The structural model is divided under three major headings: Inputs, Processes, and Outputs. Functional relationships are indicated by arrows leading from inputs through processes to outputs. In order to minimize perceptual confusion, several arrows have been omitted from the figure; these omissions will be discussed below, along with their appropriate locations and functions. The following discussion of the model is organized around the three major analyses comprising the processes by which input data are transformed into outputs. These major analyses are: Transfer Potential, Learning Deficit, and Training Techniques. Detailed discussion of the procedures involved in conducting each analysis will be presented in Section 3.0 below.

2.2.1 Transfer Potential. The model presents transfer potential as a joint function of (1) task communality between the training situation and the operational setting, and (2) similarity between the training device and the operational equipment. Task communality refers to whether a specific task in the operational situation is represented in the training device. Communality can be assessed in two ways. First, communality can be determined from the actual way that the device is currently used by the Army. For example, for the subtask, "Indexes ammunition into the computer," SIMFIRE has no communality with the operational task since trainees are presumed not to index ammunition in the SIMFIRE trainer. The other way of viewing communality is to assess "potential" communality between the device and the operational situation. In the example above, SIMFIRE would have potential communality with the operational task, since it would certainly be possible for trainees to index ammunition in the SIMFIRE trainer.

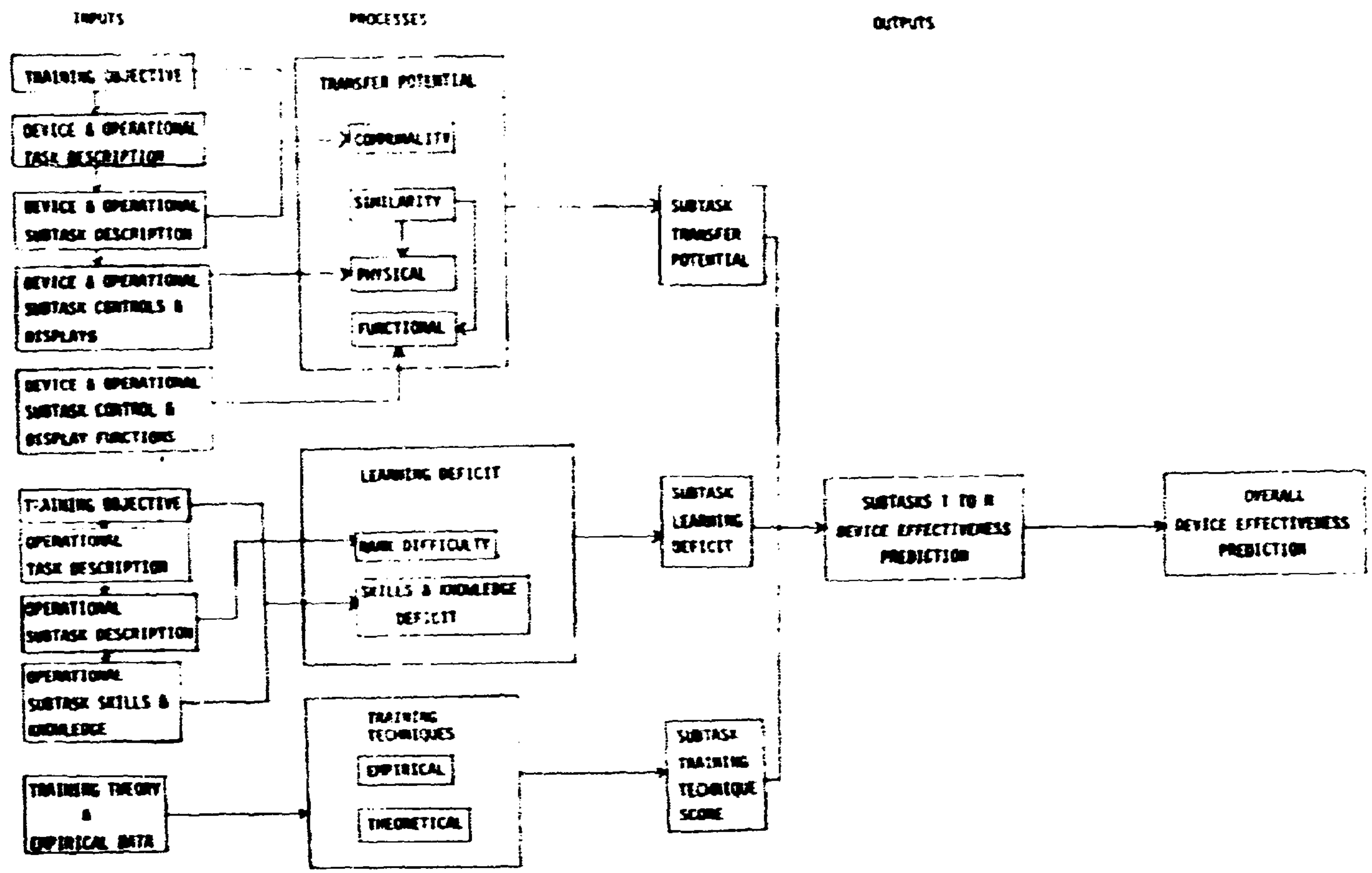

Figure 1. Structural and functional model of training device effectiveness.

Similarity refers to (1) the physical similarity between the displays and controls in the training device and those in the operational situation, and (2) the degree of functional similarity between the controls and displays in the two situations. Physical similarity is what has traditionally been referred to as "fidelity." It refers to the appearance and physical characteristics of the displays and controls in the training device and the operational situation. Functional similarity is a newer concept which refers to what the trainee actually does to the controls and displays in the two situations. The concept is based on information-processing theory, and utilizes the idea of information transmission (Fitts and Posner, 1967). Two basic decisions are made in determining functional similarity. First, are the information-processing functions of the displays and controls in the operational situation represented realistically in the training device? Function refers to such formal information-processing activities as conservation, reduction, or generation. For example, a task involving conservation is one in which the output has some fixed relationship to stimulus events, so that the stimulus (input) can be inferred precisely from the response (output). To illustrate, in the operational situation, the earphones serve the function of conservation: they transmit signals from the tank commander to the gunner regarding type of ammunition, etc. In the training situation, the earphones are not present; however, the same information conservation function is maintained since the same information is still transmitted to the gunnery trainee. The second decision is with respect to the amount of information transmitted in the two situations. For example, in the operational situation assume that the ammunition command reduces the gunner's uncertainty as to which of six types of rounds he will have to index. If, in a training device, it is possible to index only two different rounds, the amount of information in the two situations is different. The concept of functional similarity will be further elaborated in the procedures section (3.0) below.

Inputs to the Communality analysis consist of subtask descriptions for the device and the operational situation. The Physical Similarity analysis requires a list and description of the controls and displays for the device and the associated operational equipment, and the Functional Similarity analysis requires a list of control and display

functions in the two situations. Operational definitions for these inputs, along with the procedures for obtaining the information, are presented in Section 3.0. The outputs from each of these analyses are combined to arrive at a Subtask Transfer Potential value for each of the subtasks. Again, combinatorial rules will be presented in later sections of this report.

The preliminary model, in addition to proposing a Communality and Similarity analysis for Transfer Potential, also included a Criticality analysis. The basic idea was to determine which of the tasks identified by the training objective and found in the operational situation would be most critical to have represented in the training device. It became apparent, however, that no task was a priori more critical than any other for operational performance; in a sense, all subtasks are, by definition, equally important since failure to perform any one of them might result in failure to perform the entire task adequately. Criticality, therefore, is currently viewed from the perspective of the trainee: what tasks or subtasks require the most training? Thus, criticality has been subsumed under the next major analysis, Learning Deficit.

2.2.2 Learning Deficit. This analysis essentially involves determining the relationship between the level of proficiency that a trainee has prior to training and the level of proficiency that is required in order to be able to perform the operational task to criterion specifications. There are two procedures involved in this analysis: (1) a determination of the "objective" difficulty of the subtasks involved in operational performance; and (2) an assessment of the specific deficits, in terms of skill and knowledge requirements, of the trainees. The first determination involves a ranking of the relevant subtasks by difficulty, independent of the specific trainee population. The second assessment involves a rating, for each skill and knowledge requirement, of the required level of proficiency and the hypothesized level of proficiency of a given trainee population.

Inputs for these analyses are again the subtask descriptions for the operational tasks, and naturally, the training objective. In addition, a list of the skills and knowledge required for each of the subtasks serves as input. This analysis is performed independent of any particular training device; Learning Deficit depends solely on a specification of what trainees must know at the end of training. As mentioned above, "Criticality" now refers to those tasks which have the largest learning deficits; these are the tasks that are most necessary to have represented in any training device. The output of the Learning Deficit analysis is a score representing the difficulty of a subtask, weighted by the severity of the specific deficits on that subtask for a given trainee population. As suggested above, it is possible to incorporate knowledge about auxiliary classroom instruction at this stage of the overall analysis. For example, classroom instruction could provide some of the necessary knowledge that would otherwise contribute to a greater deficit rating; this information would result in a change in the weights assigned to a particular subtask.

2.2.3 Training Techniques. The third major analysis involves an assessment of the extent to which the various training steps in the device utilize various principles and techniques of training. These principles and techniques were generated from an analysis of the existing training and transfer literature. The basic procedure is to compare a specific device configuration against these training techniques and determine whether or not (and to what degree) the device conforms with or violates good practice.

The inputs to this analysis are unspecified in Figure 1. In effect, any and all information obtained from the preceding analyses is considered as input; this includes the training objective, skill and knowledge requirements, similarity ratings, and so forth. The principles themselves have been rated with respect to empirical and theoretical support.

Logistically, the Training Technique analysis is a most time-consuming activity; presently, there are approximately one hundred principles and techniques to be evaluated for each subtask, for each device under consideration. Various strategies have been employed to facilitate this analysis. One technique is to categorize each subtask and principle into an appropriate behavioral category (see Section 3.6, and Appendix A); this reduces the number of principles scanned for each subtask. Another technique is to divide the principles into categories which pertain to various aspects of performance. In the examples processed for the present report, these categories were: stimulus parameters, response parameters, and feedback parameters. Further modification and organization of this analysis is being considered.

The output of this analysis is a rating, for each subtask, of the potential effectiveness of a particular device for overcoming a specified learning deficit in a given subtask. This output is then combined with the outputs from the Transfer Potential and Learning Deficit analyses to produce a subtask device effectiveness rating. Finally, the ratings for each of the subtasks are combined to produce an overall device effectiveness rating.

Succeeding sections will further elaborate the derivation of inputs, the specific methodology for the processes, and the mathematical and practical rationale for the combinatory rules. The specific procedures have implications for informational requirements of the model as well as for experimental paradigms necessary to test it. Aside from the specific application, however, we feel that the model in its present form incorporates the various considerations applicable to the evaluation of training device effectiveness. While the adequacy of the procedures and the relations between the various outputs remain to be empirically validated, the concepts and constructs underlying the model were generated from a thorough consideration of existing literature, expert opinion, and practical guidelines of training. As such, the model can be considered apart from its specific application as the current best guess as to what should be considered in the evaluation of a device.

3.0 PROCEDURES FOR APPLICATION OF THE MODEL

3.1 Overview

In this section procedures are presented for application of the model to generate forecasts about the effectiveness of training devices. The intention is to describe the analytic steps in sufficient detail to enable others to apply the model. In the Army's case the required analyses would be performed by personnel experienced in task analysis and in derivation of training objectives. In the following portions of this section (3.2-3.6) five basic analyses are described including: Task Communality analysis (3.2), Physical Similarity analysis (3.3), Functional Similarity analysis (3.4), Learning Deficit analysis (3.5), and Training Technique assessment (3.6). In the final portion of this section (3.7), procedures are presented for determining summary index values for the three components of the structural model: Transfer Potential, Learning Deficit, and Training Techniques.

Each analysis is described according to a standard format. First, the input data requirements are indicated and the steps necessary to generate these inputs are discussed. Next, the specific procedural steps involved in conducting the analysis are described. Examples are given of how actual input data were analyzed by the project staff. The results of this application are presented along with a discussion of the feasibility of obtaining the required input data, the reliability of the procedures followed, and the problems encountered and their resolution.

In each instance when a specific analysis was attempted, the same test-bed was employed. The effectiveness of four devices in meeting two different training objectives was determined. The training devices consisted of: (1) the 17-4 Burst-on-Target Trainer; (2) the 17-B4 Burst-on-Target Trainer; (3) the M55 Conduct of Fire Trainer; and (4) SIMFIRE. The tasks underlying the two training objectives were: (1) "fire the main gun (M60A1 tank) using the M-32 sight"; and (2) "apply burst-on target adjustment of fire using the M-32 sight."

3.2 Task Communality Analysis (TCA)

As indicated above in Section 2.0, the first major analytic step in applying the model is to conduct a Task Communality analysis. The

basic purpose of this first step is to describe the overlap in training content which exists between the operational situation and any designated training device. The amount of this overlap is crucial since it is assumed that the potential for transfer of training (see Figure 1) will increase as a direct function of the degree to which criterion task relevant content is contained within the training device.

Specifically, overlap of content refers to the degree to which subtasks comprising the operational criterion situation (i.e., fire main gun using M-32 sight) are represented in the device. On this basis communality is said to exist when the trainee can perform and practice a subtask in the training situation which is also performed in the operational situation.

3.2.1 Data Requirements. The first and most basic data requirement in conducting TCA, or any of the other analyses for that matter, is a detailed statement of the training objective. The importance of satisfying this requirement cannot be emphasized strongly enough. As implied in Figure 1 above, it is the stated training objective which serves to focus attention on a specific criterion situation including: 1) the precise nature of the task to be learned; 2) the conditions of task performance during transfer; and 3) the performance standard(s) to be met. Information on each of these aspects of the stated training objective is vital for application of different portions of the model.

The procedures for developing a detailed statement of the training objective have been formalized and are presented in CON REG 350-100-1 (1972) and CON PAM 350-11 (1973). As will be discussed in a later section of the report, however, this formal detailing is not typically done during earliest stages of device development. Consequently, the statements of training objectives which accompany Training Device Requirement (TDR) documents are often fragmentary and of a very general nature. This lack of detail impacts upon application of the model early in device development.

There is a second major data requirement for TCA. Detailed task-analytic information is needed regarding the operational criterion task which has been specified in the training objective; similar data are needed regarding the training task itself. For TCA, identification and listing of the subtasks comprising the operational criterion task and the training task are essential.

In order to generate such data the general approach described in CON PAM 350-11 (1973) and the specific guidelines provided by Folley (1964) and Chenzoff and Folley (1965) may be used. To provide inputs for TCA the approach is to break the operational criterion situation down into successively finer units of description, stopping at what constitutes the subtask level of detail. Based on Folley's (1964) system of description, a subtask may be defined as an activity that is performed by one person and bounded by two events. An example of a subtask might be, "Upon receipt of the alert element of a fire command, sets turret power switch to the 'ON' position". An event may be defined as a discrete and identifiable act or occurrence. Examples would be, "receipt of alert element" and "switch in 'ON' position". An activity is defined as the behavior(s) comprising a subtask, such as "setting a switch". A task is defined as a set of two or more subtasks (e.g., "fire main gun using the M-32 sight") and, finally, a system block is defined as a set of tasks occurring at about the same time in system operation, all directed toward achieving the same sub-objective in the operation.

For the applications of the model described in the current report, task descriptive data generated by the U.S.A. Armor School, Fort Knox, Kentucky, by Powers and McCluskey (1976), and by the project team were subjected to the detailing described above. From among the 65 general tasks associated with MOS 11E (Armor Crewman), one was chosen for study: "Fire M60/M60A1 Main Gun". This system block was comprised of 15 tasks, two of which were selected for analysis. As previously indicated, these consisted of: 1) "firing the main gun with the M-32 sight"; and 2) "applying the burst-on-target (BOT) method of adjusting fire using the M-32 sight". Each task was then analyzed in order to generate a listing of

its component subtasks. The results of this analysis, shown in Table 1, served as the primary input to TCA.

3.2.2 TCA Procedure. The first step in TCA is to construct a task-communality matrix as shown in Table 2. Task and subtask information about the operational situation is listed down the left margin. The training devices to be assessed are listed across the top of the page. Notice that the listing of devices is repeated in order to permit separate analyses to be performed with respect to potential and actual communality. Potential communality addresses the overlap between subtasks in the training and operational settings which could conceivably exist, regardless of how the device is currently used. As such it represents a maximum estimate. When data are available describing how the device is actually used (i.e., which subtasks are practiced) a second estimate of communality is possible. This second estimate will be the more accurate of the two, but will usually be available only in cases where a prototype of the device is in use. (It is even conceivable that the same device, evaluated at two different locations, could produce different TCA estimates, due to differences in device utilization.) In any event, comparison of the two analyses may prove useful. When a large discrepancy exists between potential and actual communality, it may indicate the desirability of revising current methods of utilization in order to take better advantage of the device's potential for positive transfer.

The second step in TCA consists of a listing of the subtasks comprising the training task. This listing is accomplished separately for each task and device under consideration. For accuracy and to insure reliability it is suggested that this step be carried out formally. Potentially valuable information may be overlooked if one simply considers each operational subtask and makes a guess about its inclusion in the device.

Armed with lists of subtasks for the device and operational setting, one can perform the third and crucial step in TCA. For each operational

TABLE 1

SUBTASK LISTING FOR MAJOR TASKS IN THE OPERATIONAL SITUATION

TASK: Fire main gun using primary sight (M-32).

Subtasks:

1. Upon receipt of the alert element of the fire command, places turret power switch in "on" position. (Alert)

2. Upon receipt of the ammunition element of the fire command, places main gun power switch in "on" position and coaxial machine gun switch to the appropriate position. (Select Gun)

3. Index appropriate ammunition into the computer, using the ammunition selector control. (Index Computer)

4. Upon receipt of the target element of the fire command, monitors unity window for target, and, when located announces, "Identified." (Monitor)

5. Upon receiving control, operates controls to place cross hairs of sight on center of target vulnerability. (Initial Aim)

6. Tracks target. (Track)

7. Upon receipt of execution element of fire command, checks final lay of the gun and applies appropriate lead. (Final Aim)

8. Hears, "Up," announces, "On the way," pauses one second, and squeezes trigger. (Fire)

TASK: Adjust fire using Burst on Target (M-32 Sight).

Subtasks:

1. Upon firing, monitors sight for target and relays as necessary to reacquire the original sight picture. (Relay)

2. Senses the round (observes burst in relation to target, and determines new aiming point on reticle) and begins to relay. (Sense)

3. Operates controls to place new aiming point on center of target vulnerability. (Apply BOT)

4. Tracks target. (Track)

5. Hears, "Up," announces, "On the way," pauses one second, and squeezes trigger. (Fire)

TABLE 2

TASK COMMUNALITY ANALYSIS

Tasks & Subtasks	Communalities*							
	Potential Devices				Actual Devices			
	17-4	17-B4	M55	SIMFIRE	17-4	17-B4	M55	SIMFIRE
I. Fire Main Gun (M32 Sight)								
1. Alert	1	1	1	1	1	1	1	1
2. Select Gun	1	1	1	1	1	1	1	1
3. Index Computer	1	1	1	1	1	1	0	0
4. Monitor	1	1	1	1	1	1	1	1
5. Initial Aim	1	1	1	1	1	1	1	1
6. Track	0	0	1	1	0	0	1	1
7. Final Aim	1	1	1	1	1	1	1	1
8. Fire	1	1	1	1	1	1	1	1
Interjudge Agreement	1.0	1.0	1.0	1.0	.88	.88	.75	.97
II. Adjust Fire Using BOT (M32)								
1. Relay	1	1	1	1	0	0	0	0
2. Sense	1	1	1	1	1	1	1	1
3. Apply BOT	1	1	1	1	1	1	1	1
4. Track	0	0	1	1	0	0	1	1
5. Fire	1	1	1	1	1	1	1	1
Interjudge Agreement	.88	.88	1.0	1.0	.88	.88	.88	.75

* Communality estimates (1 or 0) based on a consensus among four judges; interjudge agreement calculated on the independent estimates prior to consensus.

subtask listed along the left margin of the task-communality matrix, the analyst scans his list of training subtasks. If, in fact, a device could or does enable the trainee to practice the subtask in question, a "1" is entered in the appropriate cell under the device. However, if that particular subtask is not represented, a "0" is entered. This process is continued until all operational subtasks have been evaluated.*

In making judgments about communality it is important to remember that the only consideration is whether or not the essence of the operational task is represented; in general the judgment will be based on the activity or behavior comprising the subtask. The issue of how well or how faithfully the subtask is represented is dealt with during Similarity analysis.

3.2.3 Results. The results of the Task Communality analyses are presented in Table 2 for each of the devices under examination. The task communality indices indicate that communality is quite high for all of the devices examined, and highest for those devices (e.g., M55 and SIMFIRE) which make use of the M60A1 tank itself. The 17-4 and 17-B4 devices exhibit less communality, primarily due to the fact that they do not simulate moving targets, making it impossible to train tracking (subtasks I-6, and II-4). It will also be noted that potential communality is slightly better than actual communality. This occurs for two reasons.

First, in this application it was assumed that the trainee does not index ammunition (subtask I-3) in the two tank-based devices. Second, in actual use none of the devices compels the trainee to reacquire his original sight picture after firing the main gun (subtask II-1). The results of both TCA's are retained and carried forward for use in compiling summary indices.

*In some cases there will be additional subtasks associated uniquely with a device and not found in the operational setting. These subtasks should be footnoted at the bottom of the task-communality matrix, and retained for further analysis.

3.2.4 Evaluation. In general, TCA proved to be straightforward, there being few problems in actually making the ratings. Rating ease was undoubtedly facilitated by the dichotomous nature of the decision; either the subtask was represented in the training device or it was not.

It should be emphasized, however, that successful application of the TCA procedure requires rather detailed task-descriptive statements for both the device and the operational equipment. After careful review of a number of Training Device Requirements (TDR'S) provided by the Army, the project staff was of the opinion that the level of detail currently provided in these documents would be generally inadequate for a detailed subtask communality analysis. This point is underscored by the fact that the only difficulties encountered in making the ratings on existing devices were due to incomplete information about the device on the part of the staff. Most of this difficulty was experienced by staff members who had not actually seen the device in question. These difficulties were readily cleared up by other staff members more familiar with the device. Consequently, TCA is judged feasible given that the required task-descriptive input data can be generated.

In Table 2 inter-judge agreement data are shown within each major task for the four devices. Communality judgments were obtained independently from four staff members who had advanced training and extensive experience in the behavioral sciences. The reported coefficients represent the proportion of cases in which the judges were in complete agreement. Thirty-two judgments were obtained for each device during the analysis of Task I; during analysis of Task II, 20 judgments were made.

In general, interjudge agreement was quite high. A high degree of confidence can be placed in the reliability of the procedure for performing TCA. Of the disagreements that did occur, the vast majority again arose when a judge had an erroneous impression of what a device did or did not do.

3.3 Physical Similarity Analysis (PSA)

The second major analytic step in applying the model consists of Physical Similarity analysis. This analysis is derived from several traditional fidelity-type measures, and deals in some detail with the similarity between physical characteristics of the training device and those of the operational situation. The assessment is based on the physical similarity or fidelity of displays and controls in the training device relative to those in the operational equipment.

3.3.1 Data Requirements. The detailed task descriptive data developed during TCA for both the training device and the operational equipment are used to generate a list of controls and displays relevant to each subtask. These lists constitute the basic input to PSA. A display is defined as an information source or transmitter, and a control as an information receiver which must be physically operated on. Information is defined in the information theoretic sense used by Fitts and Posner (1967). A control or display is included in the list generated for each subtask if it either transmits or receives the information involved in performance of the subtask.

In Table 3, the two major tasks together with their subordinate subtasks are listed along the left margin. Under each subtask the displays (D) and controls (C) involved directly in subtask performance are also listed. In subtask I-1, for example, information is transmitted by: 1) an earphone displaying the alert element of the fire command; and 2) by the momentary on/off position of the turret power toggle switch. The control in this subtask consists of the turret-power toggle switch itself.

3.3.2 PSA Procedure. For each subtask, a rating is made on each relevant control and display which describes how well it is represented in the training device. While ratings of subtasks lacking in communality are not used directly, it is generally useful to make ratings for all subtasks. The ratings of physical similarity are made along the following four-point scale:

TABLE 3

SIMILARITY ANALYSIS

TASK I: Fire Main Gun (M-32 Sight)	Similarity							
	Physical				Functional			
Subtasks:	17-4	17-B4	M55	SIM	17-4	17-B4	M55	SIM
1. Alert								
D_1 Earphones (gunner)	0	0	0	0	3	3	3	3
D_2 On/off pos of turret power switch	2	2	3	3	3	3	3	3
C_1 Turret power toggle switch	2	2	3	3	3	3	3	3
2. Select Gun								
D_1 Earphones (ammo)	0	0	0	0	3	3	3	3
D_2 On/off pos of Main Gun Toggle Switch	2	2	3	3	3	3	3	3
D_3 On/off pos of coaxial machine gun switch	2	2	3	3	3	3	3	3
C_1 Main gun toggle switch	2	2	3	3	3	3	3	3
C_2 Coaxial machine gun switch	2	2	3	3	3	3	3	3
3. Index Computer								
D_1 Earphones (ammo)	0	0	0	0	3	3	3	3
D_2 Indexing Window	2	2	3	3	2	2	3	3
D_3 Indexing handle feedback	1	1	3	3	1	1	3	3
C_1 Indexing handle	1	1	3	3	1	1	3	3
4. Monitor								
D_1 Earphones (target)	0	0	0	0	3	3	3	3
D_2 Unity Window	2	2	3	3	2	2	2	2
C_1 Microphone (identified)	0	0	0	0	3	3	3	3
5. Initial Aim								
D_1 Cadillac control feedback	0	0	3	3	0	0	3	3
D_2 M-32 Sight (target)	2	2	2	2	2	2	2	2
D_3 Reticle	2	3	3	3	3	3	3	3
C_1 Cadillac elevating-traversing control	1	3	3	3	2	3	3	3
C_2 Palm switches	0	3	3	3	0	3	3	3
6. Track								
D_1 M-32 Sight (dynamic target)	0	0	2	2	0	0	2	2
D_2 Reticle	2	3	3	3	0	0	3	3
C_1 Cadillac elevating-traversing control	1	3	3	3	0	0	3	3
C_2 Palm switches	0	3	3	3	0	3	3	3

TABLE 3 (Cont'd)

SIMILARITY ANALYSIS

	Similarity							
	Physical				Functional			
	17-4	17-B4	M55	SIM	17-4	17-B4	M55	SIM
7. Final Aim								
D_1 Earphones (fire)	0	0	0	0	3	3	3	3
D_2 Primary sight (target)	2	2	2	2	2	2	1	1
D_3 Reticle	2	3	3	3	3	3	3	3
C_1 Cadillac elevating-traversing control	1	3	3	3	2	3	3	3
C_2 Palm switches	0	3	3	3	0	3	3	3
8. Fire								
D_1 Earphone (up)	0	0	0	0	3	3	3	3
C_1 Microphone (on the way)	0	0	0	0	3	3	3	3
C_2 Trigger switches	2	3	3	3	3	3	3	3
TASK II: Adjust Fire Using BOT (M-32 Sight)								
Subtasks:								
1. Relay								
D_1 Fire feedback (rock-bang)	0	0	0	0	0	0	0	0
D_2 M-32 Sight (target)	2	2	2	2	2	2	2	2
D_3 Reticle	2	3	3	3	3	3	3	3
C_1 Cadillac elevating-traversing control	1	3	3	3	2	3	3	3
C_2 Palm switches	0	3	3	3	0	3	3	3
2. Sense								
D_1 M-32 sight (burst)	1	1	1	1	1	1	1	1
D_2 M-32 sight (target)	2	2	2	2	2	2	2	2
D_3 M-32 sight (reticle)	2	3	3	3	3	3	3	3
3. Apply BOT								
D_1 M-32 sight (target)	2	2	2	2	2	2	2	2
D_2 Reticle	2	3	3	3	3	3	3	3
C_1 Cadillac elevating-traversing control	1	3	3	3	2	3	3	3
C_2 Palm switches	0	3	3	3	0	3	3	3
4. Track								
D_1 M-32 sight (dynamic target)	0	0	2	2	0	0	2	2
D_2 Reticle	2	3	3	3	3	3	3	3
C_1 Cadillac elevating-traversing control	1	3	3	3	2	3	3	3

TABLE 3 (Cont'd)

SIMILARITY ANALYSIS

	Similarity							
	Physical				Functional			
	17-4	17-B4	M55	SIM	17-4	17-B4	M55	SIM
C_2 Palm switches	0	3	3	3	0	3	3	3
5. Fire								
D_1 Earphones (up)	0	0	0	0	3	3	3	3
C_1 Microphone (on the way)	0	0	0	0	3	3	3	3
C_2 Trigger switches	2	3	3	3	3	3	3	3
Initial interjudge agreement (3 of 4 agree)	.66	.76	.97	.86	.79	.86	.71	.75
Consensual interjudge agreement (4 of 4 agree)	.93	.93	.93	.93	1.0	1.0	1.0	1.0

Rating	Definition
3	Identical. The trainee would not notice a difference between the training device control or display and the operational control or display at the time of transfer. Note that they need not be absolutely identical, but there must be no "jnd" (just noticeable difference) for the trainee. Include for consideration the location, appearance, feel, and any other physical characteristics. Ignore the amount and quality of information transmitted.
2	Similar. There would be a jnd for the trainee at the time of transfer, but he would be able to perform the task. There might be a decrement in performance at transfer, but any such decrement would be readily overcome.
1	Dissimilar. There would be a large noticeable difference, quite apparent to the trainee, at transfer, and a large performance decrement, given that the trainee could perform at all. Specific instruction and practice would be required on the operational equipment after transfer to overcome the decrement.
0	The control or display is not represented at all in the training device.

The ratings are then entered in the appropriate cell of the task similarity matrix (see Table 3).

3.3.3 Results. The ratings of physical similarity for the displays and controls associated with each subtask are shown in Table 3 for the four training devices under consideration. Notice, as one might expect, that the training devices which make use of the actual operational equipment (e.g., M55 and SIMFIRE) rate quite highly in terms of physical similarity. The 17-4 and 17-B4 devices tend to have somewhat lower ratings. This is particularly true, for instance, in subtask I-1 where earphones are not used, and in subtask I-5 where cadillac control (jerk) feedback is simply not present.

Once the ratings for each subtask are completed it is possible to generate a summary index of display-control similarity at the subtask level within each device. First the mean rating for the subtask is obtained, by summing the ratings and dividing by the number of displays and controls rated for the subtask. For example, the average rating for subtask I-5 as represented in the 17-4 (see Table 3) is obtained by summing the 17-4 similarity ratings for the I-5 displays and controls (0+2+2+1+0 = 5), and dividing by the number of displays and controls rated in subtask I-5(s), providing an average rating of 5/5 = 1. This average rating will always be a positive number between 0 and 3; to provide a similarity index scale ranging from 0 to 1 the average rating is divided by 3. For subtask I-5, the 17-4 gets an index value of (5/5)/3 = .33; the same subtask receives a higher physical similarity score as represented in SIMFIRE, i.e., (14/5)/3 = .93.

3.3.4 Evaluation. The feasibility of performing PSA was related to the availability of detailed task-descriptive data on the one hand, and on the other to the analyst's familiarity with the operational equipment and training device. Conducting PSA on the basis of information contained in a TDR would be difficult unless data were available from other sources about the displays and controls comprising the operational gear. These inputs coupled with statements about the degree of realism planned for a device might permit PSA to be performed.

The same four judges who performed TCA also were involved in PSA. As shown in Table 3 reasonably high interjudge agreement was obtained for the four devices. In rating 29 different controls and displays, initial agreement among at least three out of four judges was obtained between 66% and 97% of the time. Following resolution of definitional problems the agreement among all four judges reached 93%. Most disagreements resulted from one of two sources. Analysts were occasionally misinformed about displays and controls in either the operational or training context. Additionally, there was some difficulty in distinguishing between scale values of "1" and "2". In the future it may be appropriate to collapse these two rating points.

3.4 Functional Similarity Analysis (FSA)

The next step in using the model is conducting a Functional Similarity

analysis (FSA). As indicated above, traditional fidelity measures typically concentrate on the representation of controls and displays, and ignore the behavior required of the operator using the equipment. The activities of the operator would seem to be at least as important for training as the physical characteristics of the controls and displays used to carry them out. The present analysis has been developed to assess the adequacy of representation of those activities in the training device.

The functional similarity analysis examines the operator's behavior in terms of the information flow from each display to the operator, and from the operator to each control. This examination is made in terms of the amount of information transmitted (Fitts and Posner, 1967) from each display to each control (regardless of the actual operational mode of transmission or reception), and the type of information-processing activity performed by the operator. Thus, regardless of the physical characteristics of a control or display, the issue is whether the operator acts upon the same amount of information in the same way in both the operational and training situations.

3.4.1 Data Requirements. This analysis makes use of the subtask descriptions and a list of the controls and displays in the operational situation. These inputs are used to generate a flow diagram of each subtask which indicates the type, amount, and direction of information flow for each control and display. Each situation in which a display transmits information to the operator (e.g., the operator reads the display) is defined as a stimulus function, and each situation in which the operator transmits information to a control (e.g., operates it) is termed a response function. Thus, the derived input for the Functional Similarity analysis is the list of information-processing functions indicated by the controls and displays of the operational situation.

3.4.2 FSA Procedure. In each subtask, the number of bits of information is determined for each stimulus and response situation, by estimating the number of states which the display or control may assume. The amount of information in the operational setting (H_{os}) is equal to $\log_2$ of the number of states in the stimulus or response functions under consideration. The amount of information in the training setting (H_{ts})

for each of the corresponding functions is estimated in the same way for each training device. Each stimulus and response function is then rated according to the following four-point scale:

Rating	Definition
3	Identical. $H_{ts} = H_{os}$.
2	Similar. $H_{ts} \approx H_{os}$; they are within one $\log_2$ unit of each other.
1	Dissimilar. $H_{ts} \neq H_{os}$; they are more than one $\log_2$ unit apart.
0	Missing. $H_{os} > 0$, and $H_{ts} = 0$.

In certain cases, ratings of 2 and 1 will be assigned to situations that have been purposely made unequal by the device designer in order to implement some training technique (e.g., augmented feedback or guidance). Such cases should be noted for consideration in the Training Techniques stage of the analysis. In other cases ratings of 3 will be assigned when the amount of information is the same or nearly so, but when the form of the information is radically different. For example, in the operational task the operator might index ammunition by pulling and turning the index handle. This handle could assume 6 positions; therefore, indexing ammunition is a $\log_2$ 6-bit task. In a training device, the same six alternatives might be present; however, ammunition might instead be indexed by pressing one of six buttons. The trainee might process this different information in a completely different way, or use a different strategy to deal with it. Such cases should also be noted for later consideration.

3.4.3 Results. The ratings of stimulus and response functions (denoted by the corresponding control or display) for each subtask for each of the four devices under consideration are also presented in Table 3. Notice that in some cases where the physical similarity is rated low, the functional similarity is high, while in other cases the two kinds of similarity correspond to one another quite well. The fact that there are differences suggests that this kind of analysis may be a valuable adjunct to the analysis of potential transfer.

Once again the separate ratings for each subtask may be compiled into an index which summarizes functional similarity at the subtask level within each device. This is done following the procedure already described for PSA.

3.4.4 Evaluation. The functional similarity analysis was performed independently by the same four members of the staff. Their experience suggested that FSA may be a little more difficult than the PSA, but not prohibitively so. It is crucial that the rater understand the formal concept of information transmission, and that he be able to describe explicitly the information flow situation under scrutiny. It was found that in most cases the formal step of drawing an information flow diagram was not necessary, but that on occasion it would greatly assist raters who did not have extensive experience with information theory. An understanding of types of information-processing activities (e.g., information reduction, information conservation, etc.) also aided analysts in refining their judgments.

Thus, given the list of control and display functions, the information-flow diagrams, and the detailed task descriptions, the functional similarity ratings are judged to be feasible.

The proportion of interjudge agreement is presented in Table 3 for each training device. Disagreements arose from three sources: 1) misunderstanding about the situation being rated, 2) misunderstanding about appropriate application of information measurement to the situation, and 3) an inability to discriminate between the values "2" and "1" on the scale when the amount of information was relatively large (e.g., in a real-world visual display). The four sets of ratings were resolved by consensus, and the resolved ratings are those presented.

3.5 Learning Deficit Analysis (LDA)

In order to predict the effectiveness of training devices, it was demonstrated in the first report in this series (Wheaton et al., 1976) that the analyst must not only compare the content, and physical and

functional characteristics of the operational situation and training devices, but must also consider the student to be trained. It is imperative that his capabilities be evaluated relative to the required performance criterion on the operational task. Learning Deficit Analysis (LDA) is designed to provide this assessment.

The learning deficit analysis consists of three procedures to: (1) assess the skills and knowledges in the student's repertory before training, (2) determine the skills and knowledges which he must possess at the time of transfer to the operational setting, and (3) estimate the difficulty (in terms of training time) of training the necessary skills and knowledges. The output of this stage of the analysis is a number for each subtask indicating the deficit possessed by the typical trainee, weighted by the relative difficulty (in terms of estimated training time on the operational equipment) of surmounting that deficit.

3.5.1 Data Requirements. In order to perform LDA a list of skills and knowledges necessary for the adequate performance of each subtask is generated. Beginning with the task descriptive data which provides information about actual performance of the subtask on the operational equipment, a sentence is written which describes the activity making up each subtask. From this statement, a list of skills and knowledges is developed for each subtask. The distinction between skills and knowledges is not critical in this analysis, and is made only for the convenience of the analyst. Note that the information input into this analysis (and correspondingly its output) is referenced only to the operational criterion task and to the kind of trainee expected. It is entirely independent of any particular training device, and, unlike any other portion of the model, must be performed only once for each subtask being considered, regardless of the number of devices being compared.

3.5.2 Procedure. The Learning Deficit Analysis begins with the application of a rating scale to estimate the "amount" of each skill or knowledge which the average trainee (of the type selected for course

enrollment) could be expected to have upon his first exposure to the training system or device. It is very important for the analyst to keep in mind when he makes his ratings that the trainee may have had some classroom training before his exposure to a given training device. If what the student learns in the classroom is ignored, LDA will indicate a deficit larger than may actually be the case, and the model will be making a prediction regarding the effectiveness of the training device without supporting classroom instruction. While this use of the model is permissible, the analyst will more likely be interested in the effectiveness of the device in concert with classroom support, in which case deficit to be overcome by exposure to the device will be smaller. This latter type of assessment is clearly fairer to the device, since it is obvious that some things are more effectively trained in the classroom. The following Repertory Scale (RS), adapted from Demaree (1961), is used to describe the level of each skill and knowledge in the student's repertory prior to the start of formal training:

Rating	Definition
0	No experience, training, familiarity, etc. with this skill or knowledge. Cannot perform a task requiring this skill or knowledge.
1	Has only a limited knowledge of this subject or skill. Has not actually used the information or skill. Cannot be expected to perform. Has had "orientation" only.
2	Has received a complete briefing on the subject or skill. Can use the knowledge or skill only if assisted in every step of the operation. Requires much more training and experience. Has received "familiarization" training only.
3	Understands the subject or skill to be performed. Has applied part of the knowledge or skill either on the actual job or a trainer. Has done the job enough times to make sure he can do it, although perhaps only with close supervision. Needs more practice under supervision. Has had "procedural" training.

Rating	Definition
4	Has a complete understanding of the subject or skill. Can do the task completely and accurately without supervision. Has received "skill" training.

After the analyst has assessed the level of skills and knowledges in the trainee's repertory, he proceeds to determine the required "amount" of each skill or knowledge which the trainee must possess at the close of training and time of transfer. The following Criterion Scale (CS), adapted from Demaree (1961) is used:

Rating	Definition
0	At the end of training, the trainee should have no experience or training.
1	Should have a limited knowledge of the subject or skill. Has not actually used the information. Is not expected to perform the task. Has completed "orientation" training.
2	Should have received a complete briefing on the subject or task. Is able to use the knowledge or skill only if assisted in every step of the operation. Requires much more training and experience to be able to perform the task independently. Has had "familiarization" training.
3	Should have an understanding of the subject or skill to be performed. Has applied part of the knowledge or skill on the actual job or a trainer. Has done the job enough times to make sure he can do it, although perhaps only with close supervision. Needs more practice under supervision. Has had "procedural" training.

Rating	Definition
4	Should have a complete understanding of the subject, or be highly skilled. Is able to perform the task completely, accurately, and independently. Has had "skill" training.

After knowledge and skills have been rated on both the repertory and criterion scales, the analyst calculates the learning deficit by subtracting the repertory rating (RS) from the criterion rating (CS) for the knowledge and skills underlying each subtask. Negative differences are set equal to zero, since they indicate that the trainee enters training with more proficiency than is necessary, and so has no deficit. The difference scores thus range from zero to four, the larger differences representing larger deficits. The difference scores are averaged within each subtask (collapsing across skills and knowledges in each subtask) to obtain a mean subtask deficit.

The deficit score by itself ignores the fact that some skills and knowledges are more difficult to acquire than others, and thus some subtasks may be more important to train than others. Presumably, more difficult subtasks are more critical for training, taking more time to train, and requiring more effective training. Therefore, the next step in the LDA procedure is to rank the subtasks in terms of estimated training time, assuming that only the operational equipment would be available for training. The analyst begins by seeking out the subtask whose estimated deficit would require the least training time on the operational equipment, and assigns it a Difficulty/Criticality rank of "1". The subtask requiring the next smallest amount of training time for surmounting its associated deficit is assigned a rank of "2", and so on, until all subtasks have been ranked. Next, the mean subtask deficits are multiplied by their corresponding ranks, to obtain a weighted learning deficit score. Finally, each such score is divided by 4 times the number of subtasks, to provide an index between 0 and 1 which reflects the size and importance of the deficit on each subtask relative to the other subtasks being analyzed.

3.5.3 Results. The list of skills and knowledges for each subtask is presented in Table 4 along the left margin. These are coded as "K1, K2,..." for the knowledges involved in subtask 1, "S1, S2,..." for the skills involved in subtask 1, and so on. Table 4 presents the difference between CS and RS ratings for each skill and knowledge, the mean subtask difference scores, the subtask mean rank weights, and the weighted learning deficit scores as described above.

3.5.4 Evaluation. The Learning Deficit Analysis was performed independently by four senior members of the project staff. Performed in the manner described above, it presented no significant difficulties, with the possible exception of the generation of skill-and-knowledge lists. This would not seem to be a critical problem, since slightly different lists would be expected to lead to similar mean deficit scores. The same staff members performed the analysis in an alternative way, rank ordering individual skills and knowledge, instead of subtasks, and weighting the CS-RS difference scores before collapsing within each subtask. The scores obtained in this way were highly correlated with the weighted learning deficit scores presented in Table 4. The correlations between pairs of judges ranged from .90 to .99, so that this more laborious procedure was felt to be unnecessary.

Interjudge differences in CS and RS rating scores were quite small and unsystematic, so ratings were averaged across the four judges. Interjudge agreement about subtask ranks was also quite high (correlations ranged from .85 to .99), and so ranks were averaged across judges in Table 4. Since similar high correlations on rating scores and ranks are expected in future applications, the averaging procedure (as opposed to resolution by consensus) has been tentatively adopted for this analysis.

3.6 Training Techniques Analysis (TTA)

The next step in predicting training device effectiveness is the Training Techniques Analysis. Many current devices have incorporated special features which are presumed to facilitate training beyond the level possible on the operational equipment. Given the high costs

TABLE 4

LEARNING DEFICIT ANALYSIS

		CS-RS	Mean CS-RS	Subtask Difficulty Mean Rank	Weighted Learning Deficit Score
Task I:	Repertory Item List (to be completed within 5 seconds for stationary target, 15 sec. for moving)				
X-1	Know procedure				
X-2	Know control and display locations				
X-3	Operate M-32 sight				
Subtask 1:	Alert		3.17	1.38	.14
K1:	X-1	3.50			
K2:	X-2	3.75			
S1:	Operate control blindly	2.25			
Subtask 2:	Select Gun		3.31	2.75	.28
K1:	X-1	3.50			
K2:	X-2	3.75			
K3:	Ammo-to-index value transformation	3.75			
S1:	Operate control blindly	2.25			
Subtask 3:	Index Computer		3.75	3.88	.45
K1:	X-1	3.75			
K2:	X-2	4.00			
K3:	Ammo-to-index value transformation	3.75			
S1:	Operate control blindly	3.50			
Subtask 4:	Monitor		3.00	4.88	.46
K1:	X-1	2.75			
K2:	X-2	3.75			
K3:	Target descriptors-to-target transformation	3.00			
S1:	Recognize targets visually	2.50			
Subtask 5:	Initial Aim		3.38	6.25	.66
K1:	X-1	3.25			
K2:	X-2	3.75			
K3:	"Center of vulnerability"	3.25			
S1:	X-3	3.50			
S2:	Aiming	3.00			
S3:	Detect transfer of control	3.50			

TABLE 4 (Cont'd)

LEARNING DEFICIT ANALYSIS

	CS-RS	Mean CS-RS	Subtask Difficulty Mean Rank	Weighted Learning Deficit Score
Subtask 6: Track		3.50	7.50	.82
K1: X-1	3.50			
K2: X-2	4.00			
S1: X-3	3.75			
S2: Tracking	2.75			
Subtask 7: Final Aim		3.42	7.25	.77
K1: X-1	3.50			
K2: X-2	4.00			
K3: "Aim-off"	3.50			
K4: "Lead"	3.00			
S1: X-3	3.50			
S2: Aiming	3.00			
Subtask 8: Fire		2.67	2.12	.18
K1: X-1	3.25			
K2: X-2	3.75			
S1: Estimate one second	1.00			
Task II: Repertory Item List (to be completed within 15 seconds for both moving and stationary targets)				
X-1 Know procedure				
X-2 Know control and display locations				
X-3 Operate M-32 sight				
Subtask 1: Relay		3.81	2.50	.48
K1: X-1	4.00			
K2: X-2	4.00			
S1: X-3	3.75			
S4: Aiming	3.50			
Subtask 2: Sense		3.46	3.00	.52
K1: X-1	3.75			
K2: X-2	3.75			
K3: "Sensing" vs. "non-sensing"	3.25			
K4: "Do not announce"	3.50			
S1: X-3	3.50			
S2: Sensing	3.00			

TABLE 4 (Cont'd)

LEARNING DEFICIT ANALYSIS

	CS-RS	Mean CS-RS	Subtask Difficulty Mean Rank	Weighted Learning Deficit Score
Subtask 3: Apply BOT		3.56	4.00	.71
K1: X-1	4.00			
K2: X-2	4.00			
S1: X-3	3.50			
S2: Aiming	2.75			
Subtask 4: Track		3.33	4.50	.75
K1: X-2	3.75			
S1: X-3	3.50			
S2: Tracking	2.75			
Subtask 5: Fire		2.83	1.00	.14
K1: X-1	3.75			
K2: X-2	3.75			
S1: Estimate one second	1.00			

associated with many devices it is precisely these features which justify using the devices instead of training on the operational equipment. This stage of the model attempts to assess the utility of such features, and to determine which principles of acquisition and transfer are adhered to, and which are violated, in the design of the device. Thus, given common content and good representation of the operational situation, the TTA attempts to answer the question: what incremental training value does a specific device possess over its real-world counterpart?

The Training Techniques Analysis relates the particular skills and knowledges which must be trained for each subtask to a set of principles and techniques which describe how best to train various kinds of content. The techniques have been assembled from a thorough review of the relevant literature (see Appendix A), and are organized into a taxonomic matrix. After identifying the appropriate set of techniques, the analyst makes a rating which describes the extent to which the device under examination utilizes the relevant principles/techniques in order to train a given subtask.

3.6.1 Data Requirements. The task-descriptive data and the skills-and-knowledges information from the Learning Deficit Analysis are required for the first stage of the TTA. Subtask descriptions are then assigned one or more of the following task-taxonomic labels (after U.S. Naval Training Device Center, 1972):

1. Recalling facts and principles
2. Recalling procedures
3. Non-verbal identification
4. Non-verbal detection
5. Using principles, interpreting, inferring
6. Making decisions
7. Continuous movement
8. Verbal detection and identification
9. Positioning and serial movement
10. Repetitive movement

11. Written verbalization
12. Oral verbalization
13. Other verbalization, including signs.

In toto, the required input data consist of the assigned taxonomic labels, the descriptions of the training device, any available information on the training system within which the device is embedded, and the set of training techniques.

3.6.2 TTA Procedure. The set of training techniques is organized along two independent dimensions. First, they have been coded according to the taxonomic category to which they apply. Second, within each taxonomic category, they have been further organized into techniques relevant to stimulus considerations, response considerations, or feedback considerations. Thus, by referring to the taxonomic label(s) which he has assigned to each subtask, the analyst can draw out those principles/techniques which correspond to the set of relevant behavioral categories, and sort them into three groups, stimulus, response, and feedback. With the operational task information and the training device and system information before him, he rates the training device for each relevant principle in each of the three categories. While performing the rating operation, he should pay special attention to any items from previous portions of the analysis which were "flagged" for attention at this stage (e.g., see section 3.3.2). The ratings are made from the following scale:

Rating	Definition
3	Optimal implementation of this technique; in complete accord with this principle.
2	Good implementation of this technique; in excellent accord with this principle.
1	Fair implementation of this technique; good accord with this principle.
0	This principle or technique was inapplicable or irrelevant. OR The device neither implemented this technique nor violated this principle.

Rating	Definition
-1	Mild violation of this training principle; implementation of a mildly opposing technique.
-2	Serious violation of this principle or technique.
-3	Complete violation of this principle; implementation of a strongly contraindicated technique.

For each subtask, the lowest obtained rating for each of the stimulus, response, and feedback considerations is selected, and these three ratings are then averaged to obtain the training technique score for the subtask. This score is then added to the constant 3 (to delete negative signs), and is divided by 6 to provide an index between 0 and 1 yielding the training technique score of the training device for each subtask. Note that this index is very conservative since it is quite sensitive to violations of principles/techniques. This index was considered preferable to one based on all the ratings since the negative impact of violations of principles was considered most critical in determining overall device effectiveness.

3.6.3 Results. The outcome of the TTA is shown in Table 5. The table shows the taxonomic categories assigned to each subtask, as well as the lowest ratings of the stimulus, response, and feedback components of each subtask across devices. Also shown are the transformed and averaged indices computed across devices for each subtask.

The patterns of averaged indices provide several kinds of information. For example, the lowest index value (i.e., .00) is obtained on subtasks I-6 and II-4 for the 17-4 and 17-B4 devices. The lack of a provision for the training of tracking in these two devices accounts for this rather low rating. On other subtasks, however, these devices compare favorably with the M55 or SIMFIRE devices (e.g., subtask I-4). Similarly, the SIMFIRE device receives a rating of .00 for subtask II-2 because its design limits the training which can be provided for "sensing" the round. In this case, violations of training principles were noted for the stimulus, response, and feedback aspects of the subtask.

TABLE 5

TRAINING PRINCIPLES/TECHNIQUES ANALYSIS

Behavioral Categories*	Principles/Techniques															
	Stimulus				Response				Feedback				Computed Index			
TASK I																
Subtask:	17-4	17B4	M55	SIM	17-4	17B4	M55	SIM	17-4	17B4	M55	SIM	17-4	17B4	M55	SIM
1. Alert 9	-1	1	1	1	1	1	1	1	0	1	1	1	.50	.67	.67	.67
2. Select Gun 5,9	-1	1	1	1	1	1	1	1	0	1	1	1	.50	.67	.67	.67
3. Index Computer 5, 9	-2	-2	-1	-1	-2	-2	-1	-1	-3	-3	-2	-2	.11	.11	.28	.28
4. Monitor 5	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	.44	.39	.33	.33
5. Initial Aim 4, 5, 9	-1	1	-1	-1	-2	-1	-1	-1	-1	-1	0	0	.28	.44	.39	.39
6. Track 5, 7, 9	-3	-3	-2	-2	-3	-3	2	2	-3	-3	2	2	.00	.00	.62	.62
7. Final Aim 5, 7, 9	-1	-1	-2	-2	-2	-1	-1	-1	-2	-2	-1	-1	.22	.28	.28	.28
8. Fire 4, 6	0	1	1	1	0	1	1	1	0	1	1	1	.50	.67	.67	.67
TASK II																
Subtask:																
1. Relay 5, 9	-1	-1	-1	-1	-2	-1	-1	-1	-1	-1	0	0	.28	.33	.39	.39
2. Sense 4, 5, 9	-1	-1	-1	-3	-1	-1	-1	-3	2	2	2	-3	.50	.50	.50	.00
3. Apply BOT 5, 7, 9	-1	1	-1	-1	-2	-1	-1	-1	-2	-2	-1	-1	.22	.39	.33	.33
4. Track 5, 7, 9	-3	-3	-2	-2	-3	-3	2	2	-3	-3	2	2	.00	.00	.62	.62
5. Fire 4, 6	0	1	1	1	0	1	1	1	0	1	1	1	.50	.67	.67	.67

* Categories 1, 2, and 3 were applicable to all subtasks of Tasks I and II in addition to those categories listed.

3.6.4 Evaluation. The Training Techniques Analysis was performed in two stages: the assignment of the taxonomic labels to the skill and knowledge components of subtasks, and the rating of the relevant principles/techniques. Four members of the project staff independently performed the first stage, and found it to be relatively straightforward, particularly if they were not constrained to assign only one label to each skill or knowledge. The assignment of taxonomic labels was reasonably consistent. Most interjudge disagreements resulted from incomplete understanding of the behavioral categories denoted by the taxonomic labels. It was not necessary to resolve disagreements of this sort, since the effect of disagreement here was only to increase the number of relevant principles to be noted. Since the labeling process was designed only to save the analyst's time, it is felt that use of multiple labels is satisfactory at the model's present stage of development.

The rating of the application of principles for each device was conducted by two project staff members working together, so no direct reliability data are available. Their impression was that while the assignment of ratings was time-consuming and awkward, the consensual judgments were fairly stable. A reasonable alternative to the present rating scale is to assign the zero point to a given training configuration (e.g., the operational equipment) and assign Training Technique ratings to devices relative to this standard.

This aspect of the procedures for application of the model is currently under revision. The analysis itself is viewed as high in potential for enhancing the predictive accuracy of the model.

3.7 Summary Indices

The analysis described above (e.g., TCA, PSA, FSA, LDA, and TTA) constitute a series of analytic procedures focused at the subtask level. The Army, however, will rarely be interested in predictions of transfer based on any one analysis, or for a single subtask. This section develops the procedures for collapsing across subtasks to obtain the three indices implied by the model: Transfer Potential, Learning Deficit, and Training

Techniques. The obtained indices for these three components are presently not used in calculating overall training device effectiveness, but may in fact be diagnostic of a device's particular deficiencies or assets.

The problem of collapsing the indices across subtasks and analyses is most easily viewed as a problem related to fuzzy set theory in logic and mathematics. Any system including a training device, an operational setting, and a training objective can be described in terms of a set of attributes or properties (e.g., communality, similarity, etc.), which, when taken together, contribute to transfer of training. The problem is to determine how to collapse the measures of each attribute so that they reflect the global property, "transfer of training". The solution is derived conceptually in the following sections (see Allen, 1974, for a more formal presentation of the mathematics involved).

3.7.1 Transfer Potential Index. Transfer potential is defined in the model as a joint function of communality (C) and similarity (S). This transfer potential function is characterized by the notion that C and S limit each other; i.e., similarity of non-communal subtasks does not contribute to transfer potential, nor does communality without similarity; but when a subtask is both in common with the operational situation, and possesses some similarity, transfer potential exists. Such an and relationship is described mathematically as a multiplicative operation. Thus, for any subtask i, the transfer potential of that subtask is defined as C_i times S_i. Note that this definition makes good intuitive sense as well. When both C_i and S_i are greater than zero, transfer potential is also greater than zero. If either is zero, transfer potential for that subtask is zero.

The relationship between subtasks is somewhat different. The similarity in subtask i does not contribute to transfer potential in subtask j, and the same holds for communality. Transfer potential for the two tasks combined should increase as the transfer potential for either subtask i or j increases. Or relationships are described mathematically by addition. Thus, for any two subtasks combined,

transfer potential could be defined as $(C_i \times S_i) + (C_j \times S_j)$. As described in Sections 3.2 and 3.3, C_i and S_i vary between 0 and 1. In order to scale this combined index between 0 and 1, it is then appropriate to divide the sum by 2. In general, the overall transfer potential for a set of subtasks may be defined as $\frac{\sum_{i=1}^{N} (C_i \times S_i)}{N}$, where N is the number of subtasks in the set being analyzed.

3.7.2 Learning Deficit Index. In section 3.5, an index for the learning deficit of each subtask (D) was defined as the difference between the Criterion Scale and Repertory Scale ratings, weighted by the rank difficulty of the subtask and divided by a constant to scale it between 0 and 1. Learning deficit for a task should increase when two subtasks i and j are combined, provided that D_i and D_j are greater than zero, suggesting that addition across subtasks is appropriate. Therefore, the overall learning deficit index for a set of subtasks may be defined as $\frac{\sum_{i=1}^{N} (D_i)}{N}$.

3.7.3 Training Technique Index. In section 3.6, an index (T) for the contributions of training techniques in each subtask was defined as the mean of the lowest ratings for stimulus, response, and feedback principles, transformed by constants to scale it between 0 and 1. As argued above, T for subtasks i and j should be independent. Therefore, the overall training technique index may be defined as $\frac{\sum_{i=1}^{N} (T_i)}{N}$.

Each of these indices is necessary in order to predict transfer; but none by itself is sufficient, provided the others are available (see Wheaton et al., 1974a, on the development of the structural model). What remains, therefore, is to indicate how one combines these three indices

to obtain a single index which is capable of predicting transfer. This is the subject of Section 4.0.

4.0 PREDICTION OF TRAINING DEVICE EFFECTIVENESS

4.1 Introduction

The goal of the present project is to predict training device effectiveness. The criterion selected for effectiveness is transfer of training (Wheaton, et al., 1976). This section presents the development of an equation whose parameters are the outcomes of the analyses discussed in Section 3.0. It predicts values which are components of a traditional transfer of training formula (Gagné, Foster, and Crowley, 1948).

4.2 Empirical Transfer

The formulae used to describe experimentally obtained transfer effects have been discussed in the two previous reports in this series, so only a brief review is necessary here. One of these formulae has been tentatively adopted for use with the model. This formula (Gagné, et al., 1948) expresses transfer (τ) in terms of the savings in time, trials, or errors achieved by an experimental group with pretraining, relative to a control group with no pretraining, to reach a specified criterion on the transfer task. This is formally expressed as $\tau = \frac{C-E}{C}$, where C is the number of trials or errors, or the amount of time required by the control group to reach the criterion, and E is the trials, errors, or time required by the experimental group to reach that same criterion, after some amount of pretraining. For example, suppose a control group was trained in main tank gunnery on the M60A1 tank, and required 25 hours of training to pass a gunnery proficiency test. An experimental group is pretrained on a gunnery simulator, and then is transferred to the M60A1 tank, where they require 10 hours of additional training to pass the same gunnery proficiency test. In this experiment, transfer would be calculated as follows: C=25; E=10; $\tau = \frac{25-10}{25} = .60$, indicating a savings of 60% in practice time on the operational equipment. Note that this formula does not consider the total amount of time that the experimental group spent on training, but rather the time to criterion

after transfer, so that the transfer value is always specific to a combination of the kind and amount of pretraining the experimental group received.

4.3 Predicting T

In order to predict T as defined by Gagnè, et al. (1948), it is necessary to predict the two parameters, C and E: the time, trials, or errors required by the control and experimental groups respectively to reach criterion on the transfer task. The control group is trained to criterion exclusively on the operational equipment, while the experimental group is first trained on a training device, and then trained to criterion on the operational equipment (i.e., the transfer task).

The control group may be thought of as a group which is trained on a training device having perfect communality and similarity with the transfer device, since they are, in fact, the same device. In the current form of the model, it is also assumed that the training techniques for the control group are optimal. (The implications of this assumption will be discussed later.) The level of proficiency (C) which the control group will achieve after some arbitrary amount of time is based on three components: 1) the content which is trained relative to the content which is to be tested; 2) the difficulty of acquiring the content required by the criterion; and 3) the value of the techniques employed to train the content. The content trained relative to the content to be tested is estimated by the transfer potential portion of the present model, and is expressed for any subtask i as $C_i \times S_i$. The difficulty in mastering the content required in subtask i is calculated by the learning deficit portion of the model, and is expressed as D_i. Finally, training techniques employed are described by the training techniques portion of the model, and their value is expressed as T_i. For any particular subtask, then, the amount learned by or proportion of deficit overcome by the group can be estimated by $C_i \times S_i \times D_i \times T_i$. The time, trials, or errors to a criterion on subtask i is assumed to be a linear function of $C_i \times S_i \times D_i \times T_i$. Since it was assumed that C_i, S_i, and T_i

would be 1 for all subtasks for the control group, the estimated time, trials, or errors for the total task will reduce to a linear function of $\sum_{i=1}^{N} D_i$. Thus, the model's estimate of C in the Gagné et al. formula is some linear function of $\sum_{i=1}^{N} D_i$.

The experimental group's score, E, can be considered as resulting from two components: 1) the amount of deficit overcome on the operational equipment, and 2) the amount of operational deficit overcome by exposure to the training device. The amount learned before transfer can be derived as was done above for the control group, except that C_i, S_i, and T_i do not necessarily equal 1, since the training situation for the experimental group need not be identical to the operational setting, as was the case for the control group. Therefore, the amount learned before transfer (i.e., on the training device) is estimated by $\sum_{i=1}^{N} C_i \times S_i \times D_i \times T_i$.

The amount that remains to be learned after transfer for the experiment group is then, what the total deficit was minus what was learned on the training device: $\sum_{i=1}^{N} D_i - \sum_{i=1}^{N} C_i \times S_i \times D_i \times T_i$. The time, trials, or errors to criterion can be estimated by a linear function of this expression. Thus, the model's estimate of E is some linear function of $\sum_{i=1}^{N} D_i - \sum_{i=1}^{N} C_i \times S_i \times D_i \times T_i$. Finally, the value of T can be estimated by substitution: let estimated T be written as $\hat{T}$; let estimated C = $\sum_{i=1}^{N} D_i$;

let estimated $E = \sum_{i=1}^{N} D_i - \sum_{i=1}^{N} C_i \times S_i \times D_i \times T_i$;

then

$$\hat{T} = \frac{\sum_{i=1}^{N} D_i - \sum_{i=1}^{N} D_i + \sum_{i=1}^{N} C_i \times S_i \times D_i \times T_i}{\sum_{i=1}^{N} D_i}$$

where C_i, S_i, T_i, and D_i are calculated as prescribed in Section 3.0. This formula reduces to a linear function of,

$$\hat{T} = \frac{\sum_{i=1}^{N} C_i \times S_i \times T_i \times D_i}{\sum_{i=1}^{N} D_i}.$$

This formula for $\hat{T}$ has several interesting and desirable properties, since derivations from it correspond to both empirical findings and theoretical predictions. First, as the task communality (C) increases, predicted transfer (in terms of savings) increases. Second, as similarity (S) increases, predicted transfer also increases. Third, as training techniques (T) improve, predicted transfer increases. Further, all three of these effects are moderated or weighted by the difficulty or deficit (D) of the relevant subtasks.

To illustrate this last property, consider a two-subtask case where subtask 1 is difficult, while subtask 2 is easy. Further, suppose there are two training devices designed to teach the whole task. One device has high communality, similarity, and training techniques for subtask 1, and low communality, similarity, and training techniques for subtask 2, the other device is just the opposite. Hypothetical values and calculations are presented in Table 6 for $\hat{T}$. The estimated Ts clearly favor the device which is good at training the hard task, as would

TABLE 6

ESTIMATES OF T WEIGHTED BY SUBTASK DIFFICULTY

	D_i	Training Device 1			Training Device 2		
		C_i	S_i	T_i	C_i	S_i	T_i
Subtask 1	.8 (hard)	1	1	1	0	0	0
Subtask 2	.3 (easy)	0	0	0	1	1	1
	$\sum D_i = 1.1$	$\sum C_i S_i T_i D_i = 1 \cdot 1 \cdot 1 \cdot (.8) + 0 \cdot 0 \cdot 0 \cdot (.3) = .8$			$\sum C_i S_i T_i D_i = 0 \cdot 0 \cdot 0 \cdot (.8) + 1 \cdot 1 \cdot 1 \cdot (.3) = .3$		
		$\hat{T} = \frac{.8}{\sum D_i} = \frac{.8}{1.1} = .727$			$\hat{T} = \frac{.3}{\sum D_i} = \frac{.3}{1.1} = .273$		

intuitively be expected. This weighting effect operates as well in less clear-cut cases.

One caveat needs to be sounded. It was assumed that the relationship between the amount to be learned (or remaining to be learned) and time, trials, or errors to a criterion is a nearly linear function. Such relationships in fact are known not to be, in general, linear. The relationship between time and performance has generally been found to be a monotonic increasing, negatively accelerated curve for acquisition situations (e.g., control group, or experimental group before or after transfer; c.f. Hull, 1951). Time before transfer and performance after transfer may have an even more complex relationship (c.f. Mandler, 1962). Nevertheless, it seems reasonable at the present time to characterize these relationships as approximately linear in situations of interest to Army trainers. Clearly, an important topic for further research is the relationship between amount of practice and performance in transfer/acquisition situations represented in the Army.

As mentioned earlier, there are also problems with assuming all T_is = 1 for the control group. This suggests that the operational equipment always assures the best training techniques, an assumption which contradicts the hopes of device designers who incorporate training features in devices (such as "problem freeze", augmented feedback, etc.). It might just as well have been assumed that all control T_is = .5, a neutral value, but this assumption would have been just as unfounded. What is clearly required is a training technique analysis which references the operational equipment. This does not seem too difficult to construct, and is an area for early improvement of the model.

Predicted $\mathcal{T}$s have been generated for the devices considered in Section 3.0 above, and these are presented in Table 7. $\hat{\mathcal{T}}$s are presented for each device with respect to the two major tasks selected for analysis. Further $\mathcal{T}$s have been estimated for each device with and without supporting classroom instruction for both potential and actual subtask communality. The table also presents estimated $\mathcal{T}$s for the

TABLE 7

TRAINING DEVICE EFFECTIVENESS PREDICTIONS: ESTIMATED

Training Utilization	Device			
Task I - Fire Main Gun	17-4	17-B4	M55	SIMFIRE
Potential	.1349	.2142	.3721	.3721
Potential with supporting classroom instruction	.1175	.2007	.3711	.3711
Actual	.1349	.2142	.3427	.3427
Actual with supporting classroom instruction	.1175	.2007	.3490	.3490
Task II - Adjust Fire Using BOT				
Potential	.1285	.2333	.3884	.3219
Potential with supporting classroom instruction	.1175	.1962	.3912	.3213
Actual	.1081	.1885	.3362	.2697
Actual with supporting classroom instruction	.0981	.1536	.3417	.2735
Combined (I and II)				
Potential	.1323	.2220	.3788	.3516
Potential with supporting classroom instruction	.1175	.1984	.3812	.3470
Actual	.1240	.2037	.3400	.3129
Actual with supporting classroom instruction	.1077	.1771	.3453	.3111

combined tasks. Note, however, that this combined value is not simply an average of the $\mathcal{T}$ s for the two subtasks; it is weighted by the number of subtasks, total learning deficit, and individual subtask $\mathcal{T}$ s. It is also true that [illegible] not directly comparable between tasks; primarily, this is due to the ranking procedure of the Learning Deficit analysis. While two subtasks (one in each task) might have the same "1" difficulty/criticality rank, they might not be equally easy to learn. Thus, while a higher value of $\mathcal{T}$ is interpreted as better transfer potential, this estimate is relative to the other devices under consideration for a given task.

Table 7 reveals some other interesting information as well. First, there is a large and differential impact on predicted effectiveness for the different utilizations. For example, there is no difference between actual and potential transfer predicted for the 17-4 and 17-B4, while there is a noticeable decrement for the M55 and SIMFIRE. Similarly, the effectiveness of the 17-4 and 17-B4 devices is reduced when supporting classroom instruction is provided; the predicted effectiveness generally improves with classroom support for the M55 and SIMFIRE. There are clear differences between the two tasks studied: while the M55 and SIMFIRE are equivalent for Task I, the M55 is clearly superior for Task II.

While it is important to explore the determinants of these results, any such speculations would be without empirical justification. Clearly, experimental evidence as to the relative efficacy of these devices for these hypothetical utilization situations is needed.

5.0 DISCUSSION

5.1 Composition of the Model

As presently structured, the training-effectiveness model provides for relatively comprehensive treatment of factors directly related to a training device which are known to impact upon transfer of training. There are other factors which should certainly be considered for inclusion in future versions of the model. These typically involve features of the training system external to the device itself; two of the more important of these are the amount of training/practice given and instructor-student acceptance of the device. Work on the potential impact of these kinds of training-system variables will continue.

As far as device-related variables go, one additional construct might be added immediately to the present model: a third similarity analysis which assesses the degree to which various adverse conditions, expected to impact on task performance, are simulated in the device. The data required for such an Environmental Fidelity Analysis (EFA) could be obtained by interviewing experienced operators as to the special or adverse conditions which occur from time to time and affect task performance (e.g., extreme temperature, reduced visibility, etc.). Building upon procedures suggested by Chenzoff and Folley (1965), it should be possible to estimate how severely each condition degrades performance, how likely it is to occur, and what specific subtasks are impacted upon. From descriptions of the training device, estimates could then be made regarding its capability for simulating each adverse condition.

Insofar as possible, an attempt has been made to incorporate the relevant thinking and constructs of other investigators into the model. To the extent that we have been successful in doing so, the model represents a coalescing of ideas about the nature of transfer and the factors which influence it.

5.2 Data Requirements

The initial applications of the model described in earlier sections of the report were undertaken for two reasons. First, such exercises afforded

an opportunity to identify and resolve any procedural difficulties which might arise. Second, and perhaps even more importantly, they provided for determination of the model's information requirements and for assessment of the feasibility of satisfying those requirements. At issue were the kinds of data needed and the anticipated quality of that data as a function of the type of device being evaluated and its stage of development.

5.2.1 Kinds of Input. The basic data requirements, as identified during conduct of the primary analyses (i.e., TCA, PSA, FSA, LDA, and TTA) are threefold. First, a detailed statement of the training objective(s) is mandatory for the device under investigation. Second, detailed task-analytic data of the operational task and the training situation are required. Third, estimates are needed of the capabilities and existing knowledge of the trainee population who will practice on the device.

The Army has developed detailed procedures for the specification of training objectives. As discussed in CON Reg. 350-100-1 (1972) a training objective is to contain: The action which the trainee must be able to perform, the conditions under which he is expected to perform, and the standards of performance he must reach. The action element determines which specific tasks are to be evaluated in TCA, PSA, and FSA. It also contributes to LDA. The conditions element would contribute to an Environmental Similarity Analysis and to LDA. Finally, the standards element is vital for accurate determination of the LDA.

Provision has also been made for the generation of detailed task-analytic information as described in CON Reg. 350-100-1 (1972) and CON Pam. 350-11 (1973). Particularly relevant are the data generated during training analysis which, when cast into Job-Task-Data-Card format, indicates the task, subtask, job task conditions, job task standards, skills and knowledge, and attitudes. These categories of data are directly relevant to most of the analyses required by the model. It should be noted, however, that these inputs are descriptive of the operational or transfer situation. Analogous data are required for description of the training situation. At present there appears to be no formal provision for the generation of such information. This fact, as will be discussed below, has implications for

the kinds of devices and stages of development which can be addressed by the predictive model.

The last kind of input required by the model is not available in any formalized sense, but can be obtained fairly readily. In order to specify parameter values during LDA, estimates are needed of the capabilities of the trainees who are to use a given device. It seems that the kind of input which is required can be obtained from a general consideration of trainees' backgrounds (i.e., the kind and amount of prior military training which they have received).

5.2.2 Quality of Input. The key issue in practical applications of the current model is the feasibility of acquiring the input data which are needed. As discussed above, several different kinds of input are necessary at rather detailed levels of description. Of concern, therefore, is the assumption that the quality of this input (and, consequently, of the output) may vary as a function of other factors. Two of these, considered during the applications reported above, are: 1) the system or nonsystem orientation of the device under evaluation; and 2) the stage during design and development when the evaluation is attempted.

The Army defines a training device as any three-dimensional object developed, fabricated, or procured specifically for improving the learning process. These devices are classified as system or nonsystem. System devices are those designed for use with one system or item of equipment (e.g., for the TOW missile, the M60A1 tank, the M16 rifle). Nonsystem devices are designed to support general military training and/or for use with more than one system or item of equipment (e.g., a burst-on-target trainer, or a main gun trainer). While this distinction may be valid for other reasons, it is not important within the context of the current model since it treats all devices as though they were system devices. This view stems naturally from the way in which the evaluation is conducted. The basic input always consists of the training objective, which either indicates directly, or certainly implies, the specific operational context to be considered. The model's concern with transfer of training as the measure of effectiveness forces consideration of the specific operational system

or equipment for which training is being given. For instance, when the effectiveness of the 17-4 Burst-on-Target trainer is to be estimated, the estimate must be made for some designated system. This is in no way meant to imply that the 17-4 would necessarily be judged equally effective when applied to, for instance, the M60A1 tank, the M16 rifle, or the 105mm howitzer systems. While the type of device does not influence the quality of inputs to the model, the manner in which the device is represented certainly does.

The model provides an estimate of training effectiveness based upon an analysis of both the operational context/equipment and the training situation/device. The amount and quality of the information which is available to describe either of these settings is a function of their stage of development. This point quickly becomes apparent when one considers the operational equipment and the training device at both conceptual and prototypic levels of development.

Four situations arise in which one might want to apply the model to forecast device effectiveness. In the first instance, assume that both the operational equipment and the device are at least at the prototype stage of development. In this case the quality of the information which could be generated would be at a maximum. However, the utility of the resultant forecast would be diminished by the extent to which large costs had already been incurred in producing the device prototype. In the second case, assume that the equipment was in prototype form and that the device was represented by a Training Device Requirement (TDR). Here, clearly, a forecast of effectiveness would have great utility. Unfortunately, however, the TDR's as currently written do not provide the information required for description of the device (AR 71-7, 1973). The third and fourth cases, in which either the equipment or both the equipment and device were described in preliminary document form, would prove similarly unmanageable. No single document could be found which contained all of the information required for analysis, nor would the available information be of sufficient quality.

This situation does not preclude early assessment of a device. It does suggest, however, that current TDR's would need to be reformatted if they were to serve as the single source for all necessary input data to the model. While this may represent a significant problem, it is felt that much of the needed information may be available from different sources. Aggregation of this data should be feasible.

5.3 Analytic Procedures

The procedures employed in the several analyses demanded by the model appear to be reasonably sound. Most of them represent a compromise over choice of appropriate level of analysis. By working at what seems to be an intermediate level of detail it is possible to introduce a reasonable degree of rigor and precision while avoiding the extremely time-consuming effort which has plagued other approaches.

The most difficult and lengthy processing involved occurs, in fact, during the generation of input data. The need for task-analytic information is inescapable, and one must continue to go through this complicated procedure. Within the model itself the Training Techniques analysis, as indicated in Section 3.0, is rather laborious. In the future, attempts must be made to conduct this analysis at the subtask level rather than at the level of individual knowledge and skills. Similarly, ratings should be made relative to the operational equipment which the device is designed to replace for training purposes.

5.4 Validity of the Model

Development of the predictive model is obviously intended to replace the costly approach of empirical evaluation of proposed devices and to permit evaluations earlier in the life-cycle of the device. Before reliance can be placed on the model, however, its own effectiveness must be determined. This means that its predictions must be checked against empirically obtained results. This process of verification represents validation of the model.

As indicated in the Second Interim Report (Wheaton, et al., 1974), "validation" of the model actually refers to two kinds of validity: a)

predictive validity, and b) construct validity. Predictive validity addresses the question of whether the model's output is useful in predicting the relative effectiveness of different devices. Construct validity refers to the degree to which the dimensions of the model hypothesized to influence transfer actually measure or represent what they purport to measure. In essence, construct validity addresses the theoretical structure of the model, its internal aspects rather than its output. It is clear, therefore, that predictive and construct validation serve two different purposes.

Predictive validity is assessed by the covariance between the model's output and some concurrent criterion measure of transfer of training. In perhaps the most basic case the model can make a prediction about the relative transfer of training arising from two training programs--one which employs one device and another in which instruction is provided by exposure to a different device. In this instance, the model would predict the relative effectiveness of alternative devices.

Construct validity is demonstrated by determining the relationship between specific predictor variables (e.g., various components of the model) and transfer measures. In this kind of study interest lies in determining the effects which parametric variations of independent predictor variables have on the criteria--in this case transfer-of-training measures. One can develop confidence in the value of any construct and in the construct validity of a set of measures when it is found that devices which result in good training and devices which train poorly differ in some ways, and that these differences are in accord with theoretical (model) predictions.

In the present project predictive validity is of more immediate concern than is construct validity. In essence, an attempt has already been made to build construct validity into the model. Constructs were included only after an extensive review of the literature, and only where powerful effects on transfer had already been demonstrated. The remaining construct validity questions and experimental studies must await the results from the predictive validation efforts. Consequently, the field validation effort which is being planned will focus on whether or not the model's predictions can be

corroborated by obtained data. If predictive validity is demonstrated then attention can return to the construct validity of various dimensions comprising the model.

REFERENCES

Allen, A. D. Measuring the empirical properties of sets. IEEE Transactions on Systems, Man, and Cybernetics, 1974. SMC-4(1), 66-73.

Chenzoff, A. P., & Folley, J. D., Jr. Guidelines for training situation analysis (TSA). Applied Science Associates, Inc., Valencia, Pennsylvania, 1965, AD472155.

Continental Army Command. Training. Systems Engineering of Training (Course Design). CON REG 350-100-1, April, 1972.

Continental Army Command. Training. Systems Engineering of Unit Training. CON PAM 350-11, January, 1973.

Demaree, R. G. Development of training equipment planning information. USAF ASD Technical Report 61-533, October, 1961, AD 267236.

Department of the Army. Military Training Aids and Army Training Aids Center. AR 71-7, October, 1973.

Fitts, P. M., & Posner, M. I. Human Peformance. Belmont: Brooks-Cole 1967.

Folley, J. D., Jr. Development of an improved method of task analysis and beginnings of a theory of training. Applied Sciences Associates, Inc., Valencia, Pennsylvania, 1964, AD445869.

Gagne, R. M., Foster, H., & Crowley, M. E. The measurement of transfer of training. Psychological Bulletin, 1948, 45, 97-130.

Hull, C. L. Essentials of Behavior. Yale Press: New Haven, Connecticut, 1951.

Powers, T. R., and McCluskey, M. R. (Human Resources Research Organization). Task analyses of three selected weapons systems. ARI Research Memorandum 76 . 1976.

Mandler, G. From association to structure. Psychological Review, 1962, 69, 415-427.

U.S. Naval Training Device Center. Staff study on cost and training effectiveness of proposed training systems. TAEG Report 1, U.S. Naval Training Center, Orlando, Florida, 1972.

Wheaton, G. R., Rose, A. M. Fingerman, P. W.,Korotkin, A. L., & Holding, D. H. American Institutes for Research. Evaluation of the effectiveness of training devices: Literature review and preliminary model. ARI Research Memorandum 76-6, 1976.

Wheaton, G. R., Rose, A. M., Fingerman, P. W., Korotkin, A. L., & Holding, D. H. Evaluation of the effectiveness of training devices: II. Evaluation plan for the preliminary model. Second Interim Report. U.S. Army Research Institute Contract No. DAHC 19-73-C-0049. Washington, D. C.: American Institutes for Research, 1974.

APPENDIX

APPENDIX A TRAINING TECHNIQUES AND PRINCIPLES

As mentioned in the first report in this series, the literature dealing with transfer of training is voluminous. Despite the enormous accumulation of empirical information, the data are generally still in a form which is both unsystematic and difficult to use. Very few attempts have been made to organize this literture into a usable body of knowledge. Major attempts are represented in the work of Willis and Peterson[1] and Micheli[2]. To be maximally effective, a training psychologist must be able to organize and synthesize the relevant literature in order to extrapolate to a specific training situation.

The approach employed by Willis and Peterson and modified by Micheli was to survey the literature in order to derive a set of common generalizations or principles about learning and transfer. A particular training situation could then be evaluated against these principles to estimate the efficacy of a specific training configuration. The basic organizational technique was to create a principles-by-task matrix and to generate training guidelines for each of the resultant cells. These writers employed slightly different categorical taxonomies for classification of tasks; despite these differences, both systems served primarily as aids to organizing training guidelines.

For purposes of the present model evaluation exercises, this same matrix approach was employed for the Training Techniques analysis. Each row in the matrix represented a subtask and each column represented a specific training principle/technique. The manner in which a device actually provided for training of each subtask was then contrasted with the principles/techniques. To aid in this process, each axis of the matrix was systematized. With respect to the subtask axis, task categorizations were used, based on Micheli's 13-category scheme. For the principle/techniques axis, a rough distinction among stimulus, response, and feedback principles was used.

Table A-1 presents the distribution of principles across task categories. In the present evaluation, the analysts used this table primarily as a guideline; in cases where a task categorization was doubtful, the principles for all conceivable categorizations were scanned. Table A-2 presents the set of principles employed in the present study. These principles were developed from three major sources: Willis and Peterson,

[1] Willis, M. P. & Peterson, R. O. Deriving training device implications from theory principles. Volume I: Guidelines for training device design, development, and use. U.S. Naval Training Device Center, Port Washington, New York, 1961, AD 264 364.

[2] In U.S. Naval Training Device Center. Staff study on cost and training effectiveness of proposed training systems. TAEG Report 1, U.S. Naval Training Device Center, Orlando, Florida, 1972.

Micheli, and the ongoing review of the empirical literature. Each of the principles adopted from either Willis and Peterson or Micheli was evaluated against the literature; empirical and theoretical support for each principle/technique was then ''rated.''

It should be re-emphasized that the behavioral taxonomy and principle groupings served primarily to aid the analysis in locating relevant principles. Other organizational and reference systems are currently being developed, as well as more precise statements of the principles themselves.

Table A-.

PRINCIPLES

TASK CATEGORY	STIMULUS	RESPONSE	FEEDBACK
Recall - Facts & Principles	1,2,3,4,5	1,2,3,26,31	1,2,27,29,30
Recall - Procedures	2,4,5,6,7,8,9	1,2,3,26,29,31	1,2,3,29,30
Non-Verbal Identification	10,11,12,35	4,5,26	1,4,5,27,29,30
Non-Verbal Detection	13,14	6	6,29,30
Using Principles, Interpreting, Inferring	15,16,36	7,17,18,19,29,30	7,27,29,30
Making Decisions	17,18,19,20,21,36	1,7,8,9,17,18,19,29,30	8,29,30
Continuous Movement	22,23,24,25,35,36	10,11,17,20,22,23,24, 27,28,31	9,10,15,17,19,22,23,24, 25,26,27,28,29,30
Verbal Detection and Identification	2,4,5,26	1,12	11,29,30
Positioning & Serial Movement	27,28,35,36	1,13,21,23,24,25,27, 28,31	9,12,15,16,17,18,19,20, 21,22,23,24,26,27,28, 29,30
Repetitive Movement	29,30,31,35	18,20,23,24,25,27,28, 31	9,15,17,18,19,20,21,22, 23,24,26,27,28,29,30,
Written Verbalization	32,33	1,9	13,29,30
Oral Verbalization	29	14,15,16,22	14,29,30
Other Verbalization, Including Signs	33,34	1,9,18	14,29,30

Table A-..

TRAINING PRINCIPLES/TECHNIQUES

Stimulus Considerations

1. Organize training around intrinsic cue components (key words, formuli, or key letters) within the fact or principle. Use these cue components as mediators to trigger recall of complete facts or principles.
   Empirical - good
   Theoretical - indifferent

2. Use mnemonics (associating recall of facts or principles with imagery, rhyme, rhythm, etc.).
   Empirical - good
   Theoretical - fair

3. Prevent decay of recall by increasing the meaningfulness of the material to be learned by providing organization to the related facts or principles.
   Empirical - good
   Theoretical - good

4. Prevent decay of recall by overlearning the original material.
   Empirical - good
   Theoretical - moderate

5. Prevent decay of recall by providing periodic refresher training.
   Empirical - fair
   Theoretical - excellent

6. Use mental rehearsal of sequential steps, if readily codable in symbolic form (mediators).
   Empirical - indifferent
   Theoretical - good

7. Maximize control of cues to ensure that the trainee is forming the proper associations, i.e., responding appropriately to the correct cues.
   Empirical - good
   Theoretical - excellent

8. In training for recall of lengthy or difficult procedures, develop redundant cue response patterns (via primary and supplementary stimuli) to trigger the sequentially next correct response.
   Empirical - fair
   Theoretical - good

9. Use programmed demonstration of procedures, up to but not beyond ability of student to understand procedures.
   Empirical - good
   Theoretical - good

10. Stimuli used in training should be nearly identical to job stimuli unless this fidelity increases problem difficulty in the initial phase of training to an unacceptable level.
    Empirical - good
    Theoretical - good

Table A-? (cont'd)

11. Vary ratio of relevant and irrelevant (transient) stimuli according to requirements of various stages of training. Maximize relevant cues and minimize irrelevant cues in early stages of training; use a realistic mix of relevant and irrelevant cues in final stages of training.
    Empirical - questionable
    Theoretical - good

12. Emphasize cues which elicit mediating responses, e.g., "self-instructions", "population stereotypes" and "natural associations".
    Empirical - good
    Theoretical - good

13. Transfer increases as the difference between reference and generalization stimulus decreases.
    Empirical - excellent
    Theoretical - excellent

14. Decrease signal-to-noise ratio as student achieves success at a given difficulty level.
    Empirical - good
    Theoretical - excellent

15. Emphasize the logical relationships which exist between the general principle and the specific application. The unique or special features of each application should be minimized while the common relationships to the general principle should be emphasized.
    Empirical - moderate
    Theoretical - good

16. Stimulus redundancy - apply principle in a large number of practice situations, while varying the stimulus context of repetitions.
    Empirical - fair
    Theoretical - good

17. Trainees must have access to potentially relevant data. In final stage of training, data should be limited to that expected in real world situations.
    Empirical - fair
    Theoretical - good

18. Guiding - early in training present logical implications of alternative choices.
    Empirical - fair
    Theoretical - good

19. Mediators - acquisition and use of mediators such as stereotypes or self-instructions facilitate the identification of response alternatives and the probability of success of each alternative.
    Empirical - fair
    Theoretical - good

20. Stimulus load - toward the end of training, present trainee with a realistic data processing load (realistic number of significant signals plus realistic noise in real time).
    Empirical - fair
    Theoretical - good

21. Stimulus generalization - vary the stimulus context of repetitions.
    Empirical - fair
    Theoretical - moderate

Table A-2 (cont'd)

22. Insure that the appropriate stimulus cues are available to the trainee continually during the performance of the task.
    Empirical - good
    Theoretical - moderate

23. Emphasis on prediction of future states (thinking ahead).
    Empirical - fair
    Theoretical - fair

24. Expose trainee to a wide range of task difficulty.
    Empirical - moderate
    Theoretical - good

25. In continuous control tasks, high fidelity is often required in (1) stimulus presentation, (2) operator response characteristics, and (3) dynamic system behavior, the evolving display-control relationship.
    Empirical - moderate
    Theoretical - good

26. Contiguity - the symbol and referent should be presented in close temporal contiguity.
    Empirical - good
    Theoretical - excellent

27. Cue development - emphasize the development and use of internal cues, such as mediators or kinesthetic cues.
    Empirical - moderate
    Theoretical - good

28. In training for lengthy serial movements, provision should be made for programming demonstrations of the lengthy serial or sequential performance according to the amount of demonstration which can be understood by the trainee. Continuing a demonstration beyond the "saturation point" will result in the association of responses with incorrect cues.
    Empirical - fair
    Theoretical - good

29. Early training - use models of correct performance as a basis for trainee to perceive critical cues of good form. Use models of component parts of task.
    Empirical - moderate
    Theoretical - good

30. Cue discrimination - perceive difference between correct and incorrect form.
    Empirical - good
    Theoretical - excellent

31. Later stages of training - the kinesthetic cues dominate (cues based on "muscle feel").
    Empirical - moderate
    Theoretical - questionable

32. Performance aids - especially in early phase of training use a performance aid or model, such as instructions, checklists or standard examples/formats to aid in perceiving need for and composing of messages.
    Empirical - indifferent
    Theoretical - fair

Table A-_ (cont'd)

33. "In-the-head" mediators - in later stages of instruction rely on "in-the-head" instructions, models, etc., to aid in perceiving need for and composing required messages.
    Empirical - indifferent
    Theoretical - fair

34. Static and dynamic models - in early phase of training use models, such as still and moving graphic displays (video tape recordings) to establish the characteristics of criterion performance.
    Empirical - fair
    Theoretical - fair

35. Pre-training methods need to take care not to make the S dependent upon the extra cues provided in the early stages of training and thus to hinder the changeover to more direct relations between input and output at a later stage.
    Empirical - good
    Theoretical - excellent

36. With very complex tasks, instruction in principles yields better results than laying down a detailed drill, while with simpler tasks the drill is at least equally effective.
    Empirical - good
    Theoretical - good

Table A-2 (cont'd)

Response Considerations

1. Make an overt response indicating the recall of facts and principles, enabling measurement. (Add appropriate phrase for each behavioral category.)
   Empirical - excellent
   Theoretical - excellent

2. Response Generalization - make job performance type responses (i.e., high fidelity responses later in training).
   Empirical - questionable
   Theoretical - questionable

3. Guide or prompt response, especially in the acquisition phase of training.
   Empirical - good
   Theoretical - good

4. To-be-learned response should occur as soon as relevant cues are perceived (contiguous occurrences of cues and response).
   Empirical - good
   Theoretical - excellent

5. The strength of a given response typically increases as a function of practice.
   Empirical - excellent
   Theoretical - excellent

6. To enable reinforcement of performance, the student, upon detecting a signal, should respond so that what is detected and time of detection can be recorded.
   Empirical - excellent
   Theoretical - excellent

7. Performance differences which are due to individual differences in ability tend to be magnified as a function of increasing task difficulty. Identical performance among given trainees is not necessarily indicative of identical learning, hence the need for a better measure of the extent to which trainees are profiting from the training situation.
   Empirical - good
   Theoretical - excellent

8. Apply decision making in a large number of practice situations while varying the stimulus context of repetitions.
   Empirical - fair
   Theoretical - good

9. Stress - when trainee will be required to perform under stress, use overlearning of skill to minimize effects of competing responses.
   Empirical - fair
   Theoretical - good

10. Repetition: highly skilled performance requires extensive practice.
    Empirical - excellent
    Theoretical - excellent

11. Make an objective measurement of the frequency and type of errors - changes in total error pattern - throughout the course of training. Measurement should be based on specific behavioral objectives.
    Empirical - good
    Theoretical - fair

Table A-2 (cont'd)

12. Symbol - referent associations (pairings) are especially amenable to "in-the-head" practice.
Empirical - good
Theoretical - excellent

13. Emphasize extensive motor response repetition or practice in order to (1) strengthen individual or component steps of the movement series, and (2) integrate these steps into a smooth sequence.
Empirical - excellent
Theoretical - good

14. Emphasize overt respondings in a social context. Practice to strengthen correct responses.
Empirical - indifferent
Theoretical - hunch

15. Record responses in context. Since the response is typically complex involving subtle relations among components, the technique of measurement should provide for recording the total complex in a manner that permits analysis of such subtle relationships.
Empirical - indifferent
Theoretical - excellent

16. When skill is not regularly used, prevent decay of recall by providing periodic refresher training.
Empirical - fair
Theoretical - excellent

17. The usefulness (for "lateral" transfer) of any learned capability will be increased if it is practiced in as wide a variety of situations as possible.
Empirical - good
Theoretical - excellent

18. Ensure that relevant subordinate capabilities have been thoroughly learned before calling on vertical (e.g., inclusion) transfer to aid the learning of "advanced" capabilities.
Empirical - good
Theoretical - questionable

19. Vertical transfer is enhanced by the variety of previous knowledge.
Empirical - fair
Theoretical - good

20. Where the whole task is a closely coordinated activity such as aiming a rifle or simulated flying of an aircraft, it is better to tackle the task as a whole. Any attempt to divide it up tends to destroy the proper coordination of action and subordination of individual actions to the requirements of the whole, and thus outweighs any advantage there might be in mastering different portions of the task seperately.
Empirical - good
Theoretical - fair

21. Where the task involves a series of component actions which have to be performed in the correct order but each is largely independent of the others, there seem to be advantages in practicing the different components separately.
Empirical - fair
Theoretical - good

22. Continuous practice facilitates mastery of complex, meaningful material and the establishment of coordinated rhythmic activity (within limits of fatigue).
    Empirical - good
    Theoretical - fair

23. Continuous practice seems to be preferred by older trainees.
    Empirical - moderate
    Theoretical - indifferent

24. Spaced practice is more efficient than continuous if only the actual duration of the sessions is counted and the time between sessions is ignored. When the time between sessions is included, continuous practice is usually more efficient.
    Empirical - good
    Theoretical - indifferent

25. Very brief pauses between practice sessions should be as effective as longer ones.
    Empirical - fair
    Theoretical - good

26. "Mental practice" in which the S performs a task in the imagination, can often be substituted for a substantial amount of practice involving full performance with little if any loss of effectiveness.
    Empirical - fair
    Theoretical - good

27. Relatively little learning occurs if Ss are passive spectators or even passive performers, but that they must be involved in active decisions and choices about what they are doing, and it is these that they will retain whether they are right or wrong.
    Empirical - excellent
    Theoretical - excellent

28. If two or more tasks have to be learned, it is most beneficial to begin with the one which elicits the greatest care and effort towards the attainment of a high standard of performance. However, if S was not allowed to continue to practice the more difficult task until a point of reasonable mastery, he would be left with an inadequate comprehension of the task, and transfer to a simpler task might be confused and less satisfactory than if he had tackled the easier task first.
    Empirical - good
    Theoretical - moderate

29. The more sub-tasks there are in the overall task, and the more they interact with one another, the more opportunity there will be for improvement, and therefore the longer improvement will continue.
    Empirical - good
    Theoretical - excellent

30. Transfer of skill from one task to another will depend not so much upon the extent to which methods possible for one are applied to the other, but the extent to which methods which have been selected for the one are applied to the other.
    Empirical - indifferent
    Theoretical - excellent

31. Effectiveness of spacing practice depends on what is done during the times between practice periods: (a) If they are spent in rehearsal of the material, learning will benefit, unless the task is fatiguing in which case continued practice may depress subsequent performance. (b) If time between practice periods are spent on another

Table A-2 (cont'd)

task, learning or later recall of the first task may be impaired, the degree of impairment depending on the degree of similarity between the two tasks.
Empirical - good
Theoretical - good

Table A-? (cont'd)

Feedback Considerations

1. Schedule KOR (knowledge of results) soon after response for maximum reinforcement. Error identification function of KOR is significant.
   Empirical - excellent
   Theoretical - excellent

2. Especially early in training, use KOR after the response to each step, for maximum reinforcement. Error identification function of KOR is significant. In later stages of training, step feedback is not so critical.
   Empirical - excellent (but mostly rats)
   Theoretical - excellent

3. As training progresses, gradually increase the delay in presenting KOR (present KOR in increments of 2 steps, then 3 steps, etc.) until the schedules of KOR approximates the operational setting.
   Empirical - good
   Theoretical - excellent

4. Immediate reinforcement (0.5 second delay) for non-verbal identification.
   Empirical - no data found
   Theoretical - hunch

5. KOR - automatic system performance feedback (e.g., if target is identified from partial cues, present the target with a full set of cues after identification; automate if possible).
   Empirical - indifferent
   Theoretical - good

6. Feedback omission schedule programmed according to stage of training: high feedback during initial stages, decreased to equivalent to operational setting or lower.
   Empirical - excellent
   Theoretical - excellent

7. Schedule KOR soon after response for maximum reinforcement. KOR should deal with both process and solution.
   Empirical - indifferent
   Theoretical - excellent

8. Early in training, evaluate each alternative solution as it is identified, and when a final choice among alternatives is made, evaluate the overall choice.
   Empirical - indifferent
   Theoretical - hunch

9. Shaping - reinforcement should be contingent upon characteristics of trainee's response so that by a process of 'successive approximations', the final desired proficiency is produced.
   Empirical - excellent
   Theoretical - excellent

10. Continuous KOR - because of the dynamic nature of the problem, the trainer should at times be presented with an on-going evaluation of his performance.
    Empirical - indifferent
    Theoretical - good

11. Incorrect as well as correct symbol-referent pairing can be strengthened by self-initiated "in-the-head" practice.
Empirical - moderate
Theoretical - excellent

12. Extensive response repetition (overlearning) by the trainee to take advantage of the built-in feedback properties of these types of tasks. Simple repetitive movements may be "automatically" reinforcing (kinesthetic feedback).
Empirical - excellent
Theoretical - good

13. Provide record of trainee's overt response to enable evaluation of trainee performance (i.e., must know what Ss are doing in order to provide KOR).
Empirical - excellent
Theoretical - excellent

14. Analyze oral verbalization recordings to evaluate trainer/team performance and provide KOR.
Empirical - excellent
Theoretical - excellent

15. Performance improvement in acquisition depends on knowledge of results (KOR). The rate of improvement depends upon the precision of KOR (within limits of meaningfulness).
Empirical - excellent
Theoretical - fair

16. Delay of KOR has little or no effect on acquisition (for simple motor movements only).
Empirical - excellent
Theoretical - fair

17. Increasing the post-KOR interval up to a point will improve performance level in acquisition (spacing during practice). The "point" is when recall of movement or KOR is affected.
Empirical - excellent
Theoretical - weak

18. The type of activity in the KOR delay or post-KOR delay interval does not influence acquisition (provided the intervening activity is not of the same type).
Empirical - excellent
Theoretical - weak

19. Withdrawal of KOR produces deterioration of performance when level of training is low or moderate.
Empirical - excellent
Theoretical - excellent

20. When KOR is delayed in acquisition, and S engages in deliberate verbal or motor activity during the delay interval, the effect of KOR withdrawal is poorer performance than when S rests.
Empirical - excellent
Theoretical - fair

21. When KOR is delayed in acquisition, and S rests during the delay interval, the effect on performance when KOR is withdrawn is no different than when immediate KOR is used.
Empirical - good
Theoretical - fair

Table A-1 (cont'd)

22. Activity in the post-KOR delay interval during acquisition worsens performance when KOR is withdrawn.
Empirical - excellent
Theoretical - good

23. After a relatively large amount of training, learning can continue when KOR is withdrawn.
Empirical - good
Theoretical - fair

24. If errors could be prevented in the first few trials (e.g., guidance), mastery of the task should be very much quicker.
Empirical - good
Theoretical - excellent

25. Guidance during training is beneficial when tracking movements have to be made with an incompatible control-display relationship.
Empirical - fair
Theoretical - fair

26. Guidance does not aid simple repetitive movements, but aids learning complex courses.
Empirical - good
Theoretical - good

27. The manner of conveying KOR is important: (a) Effectiveness is greatest when the information is clearly and simply related to the action performed. Any distortion or equivocation in the information fed back to the S will reduce its effectiveness. (b) Unduly full or complex information may be partly ignored or may confuse the S. (c) The information given should indicate the discrepancy between what is required and what has been achieved rather than merely give a reminder of requirements or some broad measure of achievement.
Empirical - excellent
Theoretical - excellent

28. Performance is best maintained when the conditions are such as to emphasize the need for S to observe the feel of his actions in order to relate then to their results.
Empirical - good
Theoretical - excellent

29. A S must have some cues to the results of his actions if he is to perform accurately at all, and training procedures will be effective in so far as they help him to observe and use such cues as are inherent in the task for which he is being trained. They will fail in so far as they provide him with extra cues on which he comes to rely but which are not available when he changes from training to the actual job.
Empirical - excellent
Theoretical - excellent

30. KOR acts as an incentive.
Empirical - excellent
Theoretical - good